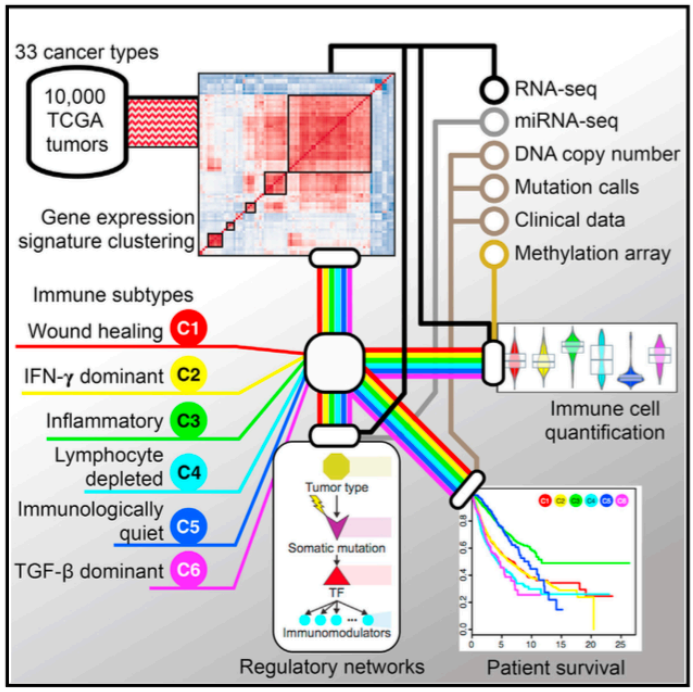
Published online on April 5, 2018

Resource

Immunity

The Immune Landscape of Cancer

Graphical Abstract



Highlights

- Six identified immune subtypes span cancer tissue types and molecular subtypes
- Immune subtypes differ by somatic aberrations, microenvironment, and survival
- Multiple control modalities of molecular networks affect tumor-immune interactions
- These analyses serve as a resource for exploring immunogenicity across cancer types

Authors

Vésteinn Thorsson, David L. Gibbs, Scott D. Brown, ..., Mary L. Disis, Benjamin G. Vincent, Ilya Shmulevich

Correspondence

vesteinn.thorsson@systemsbiology.org (V.T.), benjamin.vincent@unchealth.unc.edu (B.G.V.), ilya.shmulevich@systemsbiology.org (I.S.)

In Brief

Thorsson et al. present immunogenomics analyses of more than 10,000 tumors, identifying six immune subtypes that encompass multiple cancer types and are hypothesized to define immune response patterns impacting prognosis. This work provides a resource for understanding tumor-immune interactions, with implications for identifying ways to advance research on immunotherapy.

Thorsson et al., 2018, Immunity *48*, 1–19
April 17, 2018
https://doi.org/10.1016/j.immuni.2018.03.023

CellPress

Immunity
Resource

The Immune Landscape of Cancer

Vésteinn Thorsson,[1,36,*] David L. Gibbs,[1,35] Scott D. Brown,[2] Denise Wolf,[3] Dante S. Bortone,[4] Tai-Hsien Ou Yang,[5] Eduard Porta-Pardo,[6,7] Galen F. Gao,[8] Christopher L. Plaisier,[1,9] James A. Eddy,[10] Elad Ziv,[11] Aedin C. Culhane,[12] Evan O. Paull,[13] I.K. Ashok Sivakumar,[14] Andrew J. Gentles,[15] Raunaq Malhotra,[16] Farshad Farshidfar,[17] Antonio Colaprico,[18] Joel S. Parker,[4] Lisle E. Mose,[4] Nam Sy Vo,[19] Jianfang Liu,[20] Yuexin Liu,[19] Janet Rader,[21] Varsha Dhankani,[1] Sheila M. Reynolds,[1] Reanne Bowlby,[2] Andrea Califano,[13] Andrew D. Cherniack,[8] Dimitris Anastassiou,[5] Davide Bedognetti,[22] Arvind Rao,[19] Ken Chen,[19] Alexander Krasnitz,[23] Hai Hu,[20] Tathiane M. Malta,[24,25] Houtan Noushmehr,[24,25] Chandra Sekhar Pedamallu,[26] Susan Bullman,[26] Akinyemi I. Ojesina,[27]

(Author list continued on next page)

[1]Institute for Systems Biology, 401 Terry Ave N, Seattle, WA 98109, USA
[2]Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC V5Z 4S6, Canada
[3]University of California, San Francisco, Box 0808, 2340 Sutter Street, S433, San Francisco, CA 94115, USA
[4]Lineberger Comprehensive Cancer Center, Curriculum in Bioinformatics and Computational Biology, University of North Carolina, 125 Mason Farm Road, Chapel Hill, NC 27599-7295, USA
[5]Department of Systems Biology and Department of Electrical Engineering, Columbia University, New York, NY 10027, USA
[6]Barcelona Supercomputing Centre, c/Jordi Girona, 29, 08034 Barcelona, Spain
[7]SBP Medical Discovery Institute, La Jolla, CA 92037, USA
[8]The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02142, USA
[9]School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ 85281, USA
[10]Sage Bionetworks, 2901 Third Ave, Suite 330, Seattle, WA 98121, USA
[11]Department of Medicine, Institute for Human Genetics, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, 1450 3rd St, San Francisco, CA 94143, USA
[12]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[13]Irving Cancer Research Center, Room 913,1130 St. Nicholas Avenue, New York, NY 10032, USA
[14]Department of Computer Science, Institute for Computational Medicine; Johns Hopkins University, Baltimore, MD 21218, USA
[15]Departments of Medicine and Biomedical Data Science, Stanford University, Stanford, CA 94305, USA
[16]Seven Bridges Genomics, Cambridge, MA 02142, USA
[17]Department of Oncology, University of Calgary, Calgary, AB T2N 4N1, Canada
[18]Universite libre de Bruxelles (ULB), Computer Science Department, Faculty of Sciences, Boulevard du Triomphe - CP212, 1050 Bruxelles, Belgium
[19]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[20]Chan Soon-Shiong Institute of Molecular Medicine at Windber, Windber, PA 15963, USA
[21]Medical College of Wisconsin, 9200 Wisconsin Avenue, Milwaukee, WI 53226 USA
[22]Division of Translational Medicine, Research Branch, Sidra Medical and Research Center, PO Box 26999, Doha, Qatar
[23]Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

(Affiliations continued on next page)

SUMMARY

We performed an extensive immunogenomic analysis of more than 10,000 tumors comprising 33 diverse cancer types by utilizing data compiled by TCGA. Across cancer types, we identified six immune subtypes—wound healing, IFN-γ dominant, inflammatory, lymphocyte depleted, immunologically quiet, and TGF-β dominant—characterized by differences in macrophage or lymphocyte signatures, Th1:Th2 cell ratio, extent of intratumoral heterogeneity, aneuploidy, extent of neoantigen load, overall cell proliferation, expression of immunomodulatory genes, and prognosis. Specific driver mutations correlated with lower (*CTNNB1*, *NRAS*, or *IDH1*) or higher (*BRAF*, *TP53*, or *CASP8*) leukocyte levels across all cancers. Multiple control modalities of the intracellular and extracellular networks (transcription, microRNAs, copy number, and epigenetic processes) were involved in tumor-immune cell interactions, both across and within immune subtypes. Our immunogenomics pipeline to characterize these heterogeneous tumors and the resulting data are intended to serve as a resource for future targeted studies to further advance the field.

INTRODUCTION

The Cancer Genome Atlas (TCGA) has profoundly illuminated the genomic landscape of human malignancy. Genomic and transcriptomic data derived from bulk tumor samples have been used to study the tumor microenvironment (TME), and measures of immune infiltration define molecular subtypes of ovarian, melanoma, and pancreatic cancer (Bailey et al., 2016; The Cancer Genome Atlas Network, 2015; The Cancer Genome Atlas

Andrew Lamb,[10] Wanding Zhou,[28] Hui Shen,[28] Toni K. Choueiri,[26] John N. Weinstein,[19] Justin Guinney,[10] Joel Saltz,[29] Robert A. Holt,[2] Charles E. Rabkin,[30] The Cancer Genome Atlas Research Network, Alexander J. Lazar,[31] Jonathan S. Serody,[32] Elizabeth G. Demicco,[33,35] Mary L. Disis,[34,35] Benjamin G. Vincent,[4,*] and Ilya Shmulevich[1,*]
[24]Department of Neurosurgery, Henry Ford Hospital, Detroit, MI 48202, USA
[25]Department of Genetics, Ribeirao Preto Medical School, University of São Paulo, São Paulo, Brazil
[26]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[27]University of Alabama at Birmingham, Birmingham, AL 35294, USA
[28]Center for Epigenetics, Van Andel Research Institute, Grand Rapids, MI 49503, USA
[29]Department of Biomedical Informatics, Stony Brook Medicine, 100 Nicolls Rd, Stony Brook, NY 11794, USA
[30]Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Dr., Bethesda, MD 20892, USA
[31]Departments of Pathology, Genomics Medicine and Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd-Unit 85, Houston, TX 77030, USA
[32]Department of Medicine and Microbiology and Lineberger Comprehensive Cancer Center, 125 Mason Farm Road, Chapel Hill, NC 27599-7295, USA
[33]Mount Sinai Hospital, Department of Pathology and Laboratory Medicine, 600 University Ave., Toronto, ON M5G 1X5, Canada
[34]UW Medicine Cancer Vaccine Institute, 850 Republican Street, Brotman Building, 2nd Floor, Room 221, Box 358050, University of Washington, Seattle, WA 98109-4714, USA
[35]These authors contributed equally
[36]Lead Author
*Correspondence: vesteinn.thorsson@systemsbiology.org (V.T.), benjamin.vincent@unchealth.unc.edu (B.G.V.), ilya.shmulevich@systemsbiology.org (I.S.)
https://doi.org/10.1016/j.immuni.2018.03.023

Research Network, 2011) and immune gene expression in other tumors varies by molecular subtype (Iglesia et al., 2016). Characterization of the immune microenvironment using gene expression signatures, T cell receptor (TCR), and B cell receptor (BCR) repertoire, and analyses to identify neo-antigenic immune targets provide a wealth of information in many cancer types and have prognostic value (Bindea et al., 2013; Brown et al., 2014, 2015; Charoentong et al., 2017; Gentles et al., 2015; Iglesia et al., 2016; Li et al., 2016; Porta-Pardo and Godzik, 2016; Rooney et al., 2015).

Contemporaneous with the work of TCGA, cancer immunotherapy has revolutionized cancer care. Antibodies against CTLA-4, PD-1, and PD-L1 are effective in treating a variety of malignancies. However, the biology of the immune microenvironment driving these responses is incompletely understood (Hugo et al., 2016; McGranahan et al., 2016) but is critical to the design of immunotherapy treatment strategies.

We integrated major immunogenomics methods to characterize the immune tumor microenvironment (TME) across 33 cancers analyzed by TCGA, applying methods for the assessment of total lymphocytic infiltrate (from genomic and H&E image data), immune cell fractions from deconvolution analysis of mRNA-seq data, immune gene expression signatures, neoantigen prediction, TCR and BCR repertoire inference, viral RNA expression, and somatic DNA alterations (Table S1). Transcriptional regulatory networks and extracellular communication networks that may govern the TME were found, as were possible germline determinants of TME features, and prognostic models were developed.

Through this approach, we identified and characterized six immune subtypes spanning multiple tumor types, with potential therapeutic and prognostic implications for cancer management. All data and results are provided in Supplemental Tables, at the NCI Genomic Data Commons (GDC, https://portal.gdc.cancer.gov), and though the Cancer Research Institute *iAtlas* portal for interactive exploration and visualization (http://www.cri-iatlas.org), and are intended to serve as a resource for future studies in the field of immunogenomics.

RESULTS

Analytic Pipeline

To characterize the immune response to cancer in all TCGA tumor samples, identify common immune subtypes, and evaluate whether tumor-extrinsic features can predict outcomes, we analyzed the TME across the landscape of all TCGA tumor samples. First, source datasets from all 33 TCGA cancer types and six molecular platforms (mRNA, microRNA, and exome sequencing; DNA methylation-, copy number-, and reverse-phase protein arrays) were harmonized by the PanCanAtlas consortium for uniform quality control, batch effect correction, normalization, mutation calling, and curation of survival data (Ellrott et al., 2018; Liu et al., 2018). We then performed a series of analyses, which we summarize here and describe in detail in the ensuing manuscript sections as noted within parentheses. We first compiled published tumor immune expression signatures and scored these across all non-hematologic TCGA cancer types. Meta-analysis of subsequent cluster analysis identified characteristic immunooncologic gene signatures, which were then used to cluster TCGA tumor types into six groups, or subtypes (described in Immune Subtypes in Cancer). Leukocyte proportion and cell type were then defined from DNA methylation, mRNA, and image analysis (see Composition of the Tumor Immune Infiltrate). Survival modeling was performed to assess how immune subtypes associate with patient prognosis (see Prognostic Associations of Tumor Immune Response Measures). Neoantigen prediction and viral RNA expression (see Survey of Immunogenicity), TCR and BCR repertoire inference (see The Adaptive Immune Receptor Repertoire in Cancer), and immunomodulator (IM) expression and regulation (see Regulation of Immunomodulators) were characterized in the context of TCGA tumor types, TCGA-defined molecular subtypes, and these six immune subtypes, so as to assess the relationship between factors affecting immunogenicity and immune infiltrate. In order to assess the degree to which specific underlying somatic alterations (pathways, copy-number

alterations, and driver mutations) may drive the composition of the TME, we identified which alterations correlate with modified immune infiltrate (see Immune Response Correlates of Somatic Variation). We likewise asked whether gender and ancestry predispose individuals to particular tumor immune responses (see Immune Response Correlates of Demographic and Germline Variation). Finally, we sought to identify the underlying intracellular regulatory networks governing the immune response to tumors, as well as the extracellular communication networks involved in establishing the particular immune milieu of the TME (see Networks Modulating Tumoral Immune Response).

Immune Subtypes in Cancer

To characterize intratumoral immune states, we scored 160 immune expression signatures and used cluster analysis to identify modules of immune signature sets (Figure 1A, top). Five immune expression signatures—macrophages/monocytes (Beck et al., 2009), overall lymphocyte infiltration (dominated by T and B cells) (Calabro et al., 2009), TGF-β response (Teschendorff et al., 2010), IFN-γ response (Wolf et al., 2014), and wound healing (Chang et al., 2004)—which robustly reproduced co-clustering of these immune signature sets, were selected to perform cluster analysis of all 30 non-hematologic cancer types (Figures 1A middle, and S1A). The six resulting clusters "Immune Subtypes," C1–C6 (with 2,416, 2,591, 2,397, 1,157, 385, and 180 cases, respectively) were characterized by a distinct distribution of scores over the five representative signatures (Figure 1A, bottom) and showed distinct immune signatures based on the dominant sample characteristics of their tumor samples (Figures 1B and 1C). Immune subtypes spanned anatomical location and tumor type, while individual tumor types and TCGA subtypes (Figures 1D and S1B–S1D) varied substantially in their proportion of immune subtypes.

C1 (wound healing) had elevated expression of angiogenic genes, a high proliferation rate (Figure 1C), and a Th2 cell bias to the adaptive immune infiltrate. Colorectal cancer (COAD [colon adenocarcinoma], READ [rectum adenocarcinoma]) and lung squamous cell carcinoma (LUSC) were rich in C1, as were breast invasive carcinoma (BRCA) luminal A (Figures S1C and S1D), head and neck squamous cell carcinoma (HNSC) classical, and the chromosomally unstable (CIN) gastrointestinal subtype.

C2 (IFN-γ dominant) had the highest M1/M2 macrophage polarization (Figure S2A, mean ratio = 0.52, $p < 10^{-149}$, Wilcoxon test relative to next-highest), a strong CD8 signal and, together with C6, the greatest TCR diversity. C2 also showed a high proliferation rate, which may override an evolving type I immune response, and was comprised of highly mutated BRCA, gastric, ovarian (OV), HNSC, and cervical tumors (CESC).

C3 (inflammatory) was defined by elevated Th17 and Th1 genes (Figure 1C, both $p < 10^{-23}$), low to moderate tumor cell proliferation, and, along with C5, lower levels of aneuploidy and overall somatic copy number alterations than the other subtypes. C3 was enriched in most kidney, prostate adenocarcinoma (PRAD), pancreatic adenocarcinoma (PAAD), and papillary thyroid carcinomas (THCA).

C4 (lymphocyte depleted) was enriched in particular subtypes of adrenocortical carcinoma (ACC), pheochromocytoma and paraganglioma (PCPG), liver hepatocellular carcinoma (LIHC), and gliomas, and displayed a more prominent macrophage signature (Figure 2A), with Th1 suppressed and a high M2 response (Figure S2A).

C5 (immunologically quiet), consisted mostly of brain lower-grade gliomas (LGG) (Figures 1D and S1B), exhibited the lowest lymphocyte ($p < 10^{-17}$) and highest macrophage ($p < 10^{-7}$) responses (Figure 2A), dominated by M2 macrophages (Figure S2A). Glioma subtypes (Ceccarelli et al., 2016) CpG island methylator phenotype-high (CIMP-H), the 1p/19q codeletion subtype and pilocytic astrocytoma-like (PA-like) were prevalent in C5, with remaining subtypes enriched in C4. *IDH* mutations were enriched in C5 over C4 (80% of *IDH* mutations, $p < 2 \times 10^{-16}$, Fisher's exact test), suggesting an association of *IDH* mutations with favorable immune composition. Indeed, *IDH* mutations associate with TME composition (Venteicher et al., 2017) and decrease leukocyte chemotaxis, leading to fewer tumor-associated immune cells and better outcome (Amankulor et al., 2017).

Finally, C6 (TGF-β dominant), which was a small group of mixed tumors not dominant in any one TCGA subtype, displayed the highest TGF-β signature ($p < 10^{-34}$) and a high lymphocytic infiltrate with an even distribution of type I and type II T cells.

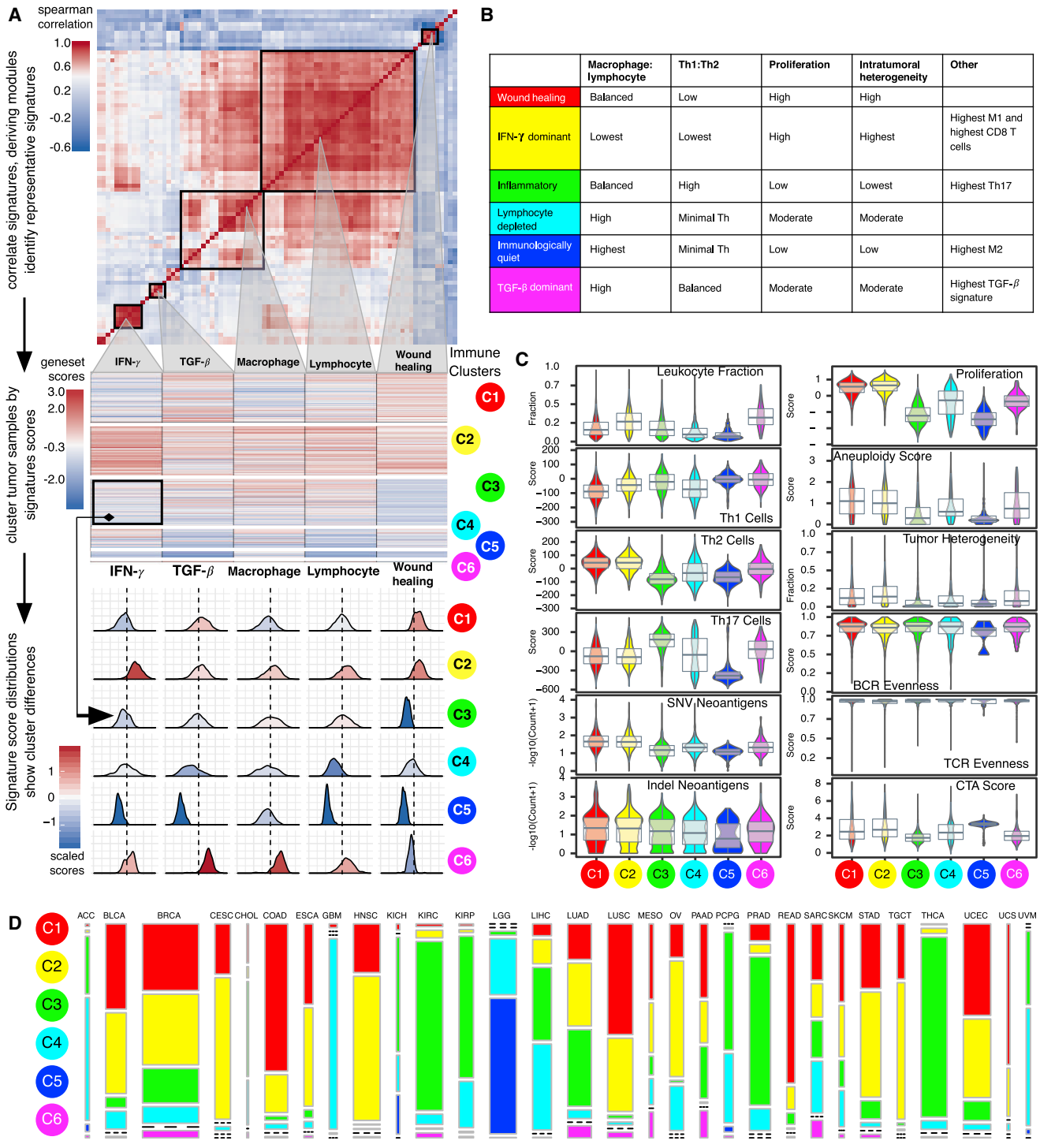
These six categories represent features of the TME that largely cut across traditional cancer classifications to create groupings and suggest certain treatment approaches may be independent of histologic type. For a complete list of the TCGA cancer type abbreviations, please see https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations.

Composition of the Tumor Immune Infiltrate

Leukocyte fraction (LF) varied substantially across immune subtypes (Figure 1C) and tumor types (Figure 2B). Tumors within the top third LF included cancers most responsive to immune checkpoint inhibitors, such as lung adenocarcinoma (LUAD), LUSC, cutaneous melanoma (SKCM), HNSC, and kidney renal clear cell carcinoma (KIRC), and in particular, the LUSC.secretory, LUAD.6, bladder urothelial carcinoma (BLCA.4), kidney renal papillary cell carcinoma (KIRP.C2a), and HNSCC mesenchymal subtypes. Uveal melanoma (UVM) and ACC had very low LF. Glioma subtypes displayed a greater range in LF than other tumors, which may reflect the presence or absence of microglia.

The leukocyte proportion of tumor stromal fraction, ρ, varied across tumor types and immune subtypes (Figures 2C and S2B), ranging from >90% in SKCM to <10% in stroma-rich tumors such as PAAD, PRAD, and LGG. Some tumors, e.g., BRCA, showed variation within annotated or immune subtypes. In BRCA, C1 has the lowest ρ ($\rho^{C1} = 0.44$) while $\rho^{C2} = 0.61$ was 37% higher ($p < 0.001$) (Figure S2B), and there were likewise differences between luminal A and basal BRCA ($\rho^{LumA} = 0.45$ and $\rho^{Basal} = 0.67$ [$p < 0.001$]). For LGG, $\rho^{C5} = 0.28$ ($p < 0.001$), whereas $\rho^{C3} = 0.48$ and $\rho^{C4} = 0.50$ ($p < 0.001$) (Figure S2B), and in READ, $\rho^{CIN} = 0.40$ and $\rho^{MSI} = 0.78$ ($p < 0.001$).

The spatial fraction of tumor regions with tumor-infiltrating lymphocytes (TILs), estimated by analysis of digitized TCGA H&E-stained slides (Saltz et al., 2018), varied by immune subtype, with C2 the highest ($p < 10^{-16}$, Figure 2D). Image estimates correlated modestly with molecular estimates of lymphocyte proportion (Figures S2C and S2D), in part because the molecular



	Macrophage: lymphocyte	Th1:Th2	Proliferation	Intratumoral heterogeneity	Other
Wound healing	Balanced	Low	High	High	
IFN-γ dominant	Lowest	Lowest	High	Highest	Highest M1 and highest CD8 T cells
Inflammatory	Balanced	High	Low	Lowest	Highest Th17
Lymphocyte depleted	High	Minimal Th	Moderate	Moderate	
Immunologically quiet	Highest	Minimal Th	Low	Low	Highest M2
TGF-β dominant	High	Balanced	Moderate	Moderate	Highest TGF-β signature

Figure 1. Immune Subtypes in Cancer

(A) Expression signature modules and identification of immune subtypes. Top: Consensus clustering of the pairwise correlation of cancer immune gene expression signature scores (rows and columns). Five modules of shared associations are indicated by boxes. Middle: Representative gene expression signatures from each module (columns), which robustly reproduced module clustering, were used to cluster TGCA tumor samples (rows), resulting in six immune subtypes C1–C6 (colored circles). Bottom: Distributions of signature scores within the six subtypes (rows), with dashed line indicating the median.

(B) Key characteristics of immune subtypes.

(C) Values of key immune characteristics by immune subtype.

(D) Distribution of immune subtypes within TCGA tumors. The proportion of samples belonging to each immune subtype is shown, with colors as in (A). Bar width reflects the number of tumor samples.

See also Figure S1 and Table S1.

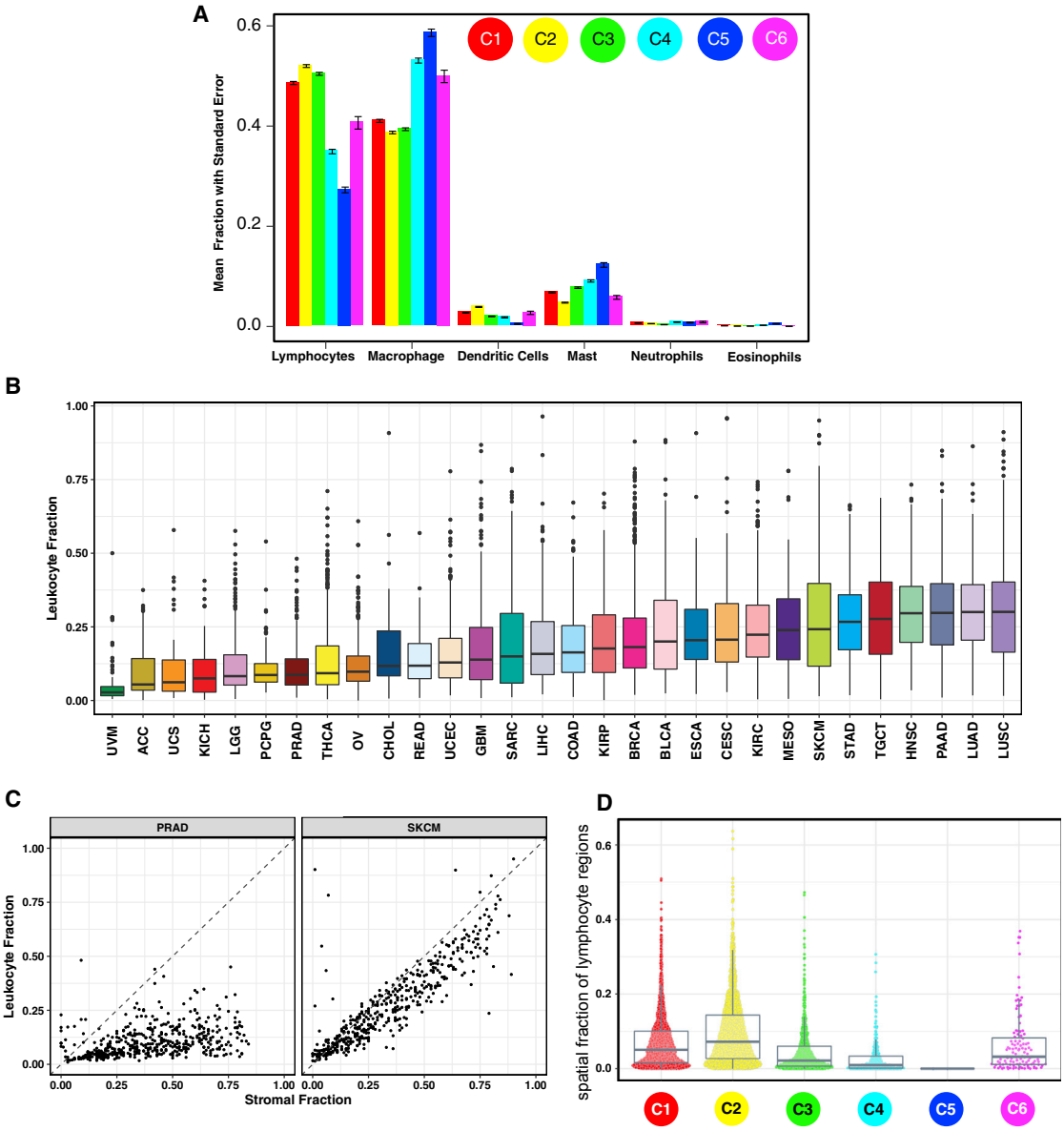


Figure 2. Composition of the Tumor Immune Infiltrate

(A) The proportion of major classes of immune cells (from CIBERSORT) within the leukocyte compartment for different immune subtypes. Error bars show the standard error of the mean.

(B) Leukocyte fraction (LF) within TCGA tumor types, ordered by median.

(C) LF (y axis) versus non-tumor stromal cellular fraction in the TME (x axes) for two representative TCGA tumor types: PRAD, (low LF relative to stromal content) and SKCM (high leukocyte fraction in the stroma). Dots represent individual tumor samples.

(D) The spatial fraction of lymphocyte regions in tissue was estimated using machine learning on digital pathology H&E images (see also Saltz et al., 2018).

estimate is more similar to cell count, while spatial TIL is a fraction of the area. The relative similarity of the estimates of lymphocytic content between two radically different methodologies reinforces the robustness of individual methods.

Prognostic Associations of Tumor Immune Response Measures

Immune subtypes associated with overall survival (OS) and progression-free interval (PFI) (Figures 3A and S3A). C3 had the best prognosis (OS HR 0.628, $p = 2.34 \times 10^{-8}$ relative to C1, adjusted for tumor type), while C2 and C1 had less favorable outcomes despite having a substantial immune component. The more mixed-signature subtypes, C4 and C6, had the least favorable outcome. Functional orientation of the TME for tumor and immune subtypes was measured using the concordance index (CI) (Pencina and D'Agostino, 2004) and found to have context-dependent prognostic impact (Figures 3B, 3C and S3B). Higher lymphocyte signature associated with improved outcome in C1 and C2. An increased value of any of the five signatures led to worse outcome in C3 (Figure 3B), perhaps

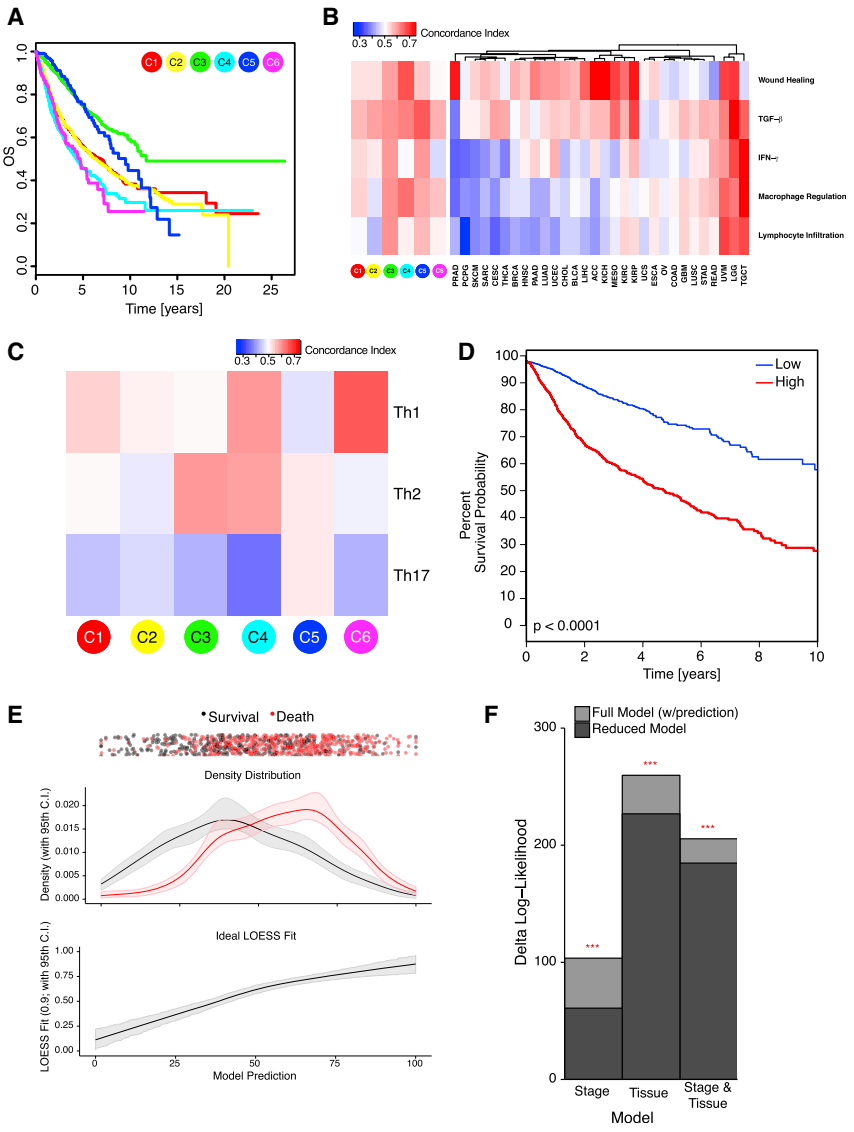


Figure 3. Immune Response and Prognostics

(A) Overall survival (OS) by immune subtype.

(B) Concordance index (CI) for five characteristic immune expression signature scores (Figure 1A) in relation to OS, for immune subtypes and TCGA tumor types. Red denotes higher and blue lower risk, with an increase in the signature score.

(C) CI for T helper cell scores in relation to OS within immune subtypes.

(D) Risk stratification from elastic net modeling of immune features. Tumor samples were divided into discovery and validation sets, and an elastic net model was optimized on the discovery set using immune gene signatures, TCR/BCR richness, and neoantigen counts. Kaplan-Meier plot shows the high (red) and low (blue) risk groups from this model as applied to the validation set, $p < 0.0001$ (*G-rho* family of tests, Harrington and Fleming).

(E) Prediction versus outcome from elastic net model in validation set data (from D). Top: Patient outcomes for each sample (black, survival; red, death) plotted with vertical jitter, along the sample's model prediction (x axis). Middle: Fractional density of the outcomes plotted against their model predictions. Confidence intervals were generated by bootstrapping with replacement. Bottom: LOESS fit of the actual outcomes against the model predictions; narrow confidence bands confirm good prediction accuracy.

(F) CoxPH models of stage and tumor type ("Tissue") with (full model) or without (reduced model) the validation set predictions of the elastic net model were compared; the full model significantly outperformed the reduced model in all comparisons ($p < 0.001$; false discovery rate (FDR) BH-corrected). See also Figure S3.

reflecting a balanced immune response. While increased Th17 cells generally led to improved OS, Th1 associated with worse OS across most immune subtypes, and Th2 orientation had mixed effects (Figure 3C). Tumor types displayed two behaviors relative to immune orientation (Figures 3B, OS; S3B, PFI). In the first group including SKCM and CESC, activation of immune pathways was generally associated with better outcome, while in the other, the opposite was seen. The relative abundance of individual immune cell types had complex associations that differed between tumor types (Figures S3C and S3D). These analyses extend beyond mere determination of lymphocyte presence to suggest testable properties that correlate with patient outcome in different tumor types and immune contexts.

We obtained and validated a survival model using elastic-net Cox proportional hazards (CoxPH) modeling with cross-validation. Low- and high-score tumors displayed significant survival differences in the validation set (Figure 3D), with good prediction accuracy (Figure 3E). Incorporating immune features into Cox models fit with tumor type, stage, and tumor type + stage (Figure 3F) improved predictive accuracy, highlighting the importance of the immune TME in determining survival. Lymphocyte expression signature, high number of unique TCR clonotypes, cytokines made by activated Th1 and Th17 cells, and M1 macrophages most strongly associated with improved OS (Figure S3E), while wound healing, macrophage regulation, and TGF-β associated with worse OS, recapitulating survival associations in immune subtypes. Within tumor types, the prognostic implications of immune subtypes seen in univariate analyses were largely maintained, with C3 correlating with better OS in six tumor types and C4 with poor OS in three cancer types (Figure S3F).

Immune Response Correlates of Somatic Variation

The immune infiltrate was related to measures of DNA damage, including copy number variation (CNV) burden (both in terms of number of segments and fraction of genome alterations), aneuploidy, loss of heterozygosity (LOH), homologous recombination deficiency (HRD), and intratumor heterogeneity (ITH) (Figure 4A). LF correlated negatively with CNV segment burden, with strongest correlation in C6 and C2, and positively with aneuploidy, LOH, HRD, and mutation load, particularly in C3. These results suggest a differential effect of multiple smaller,

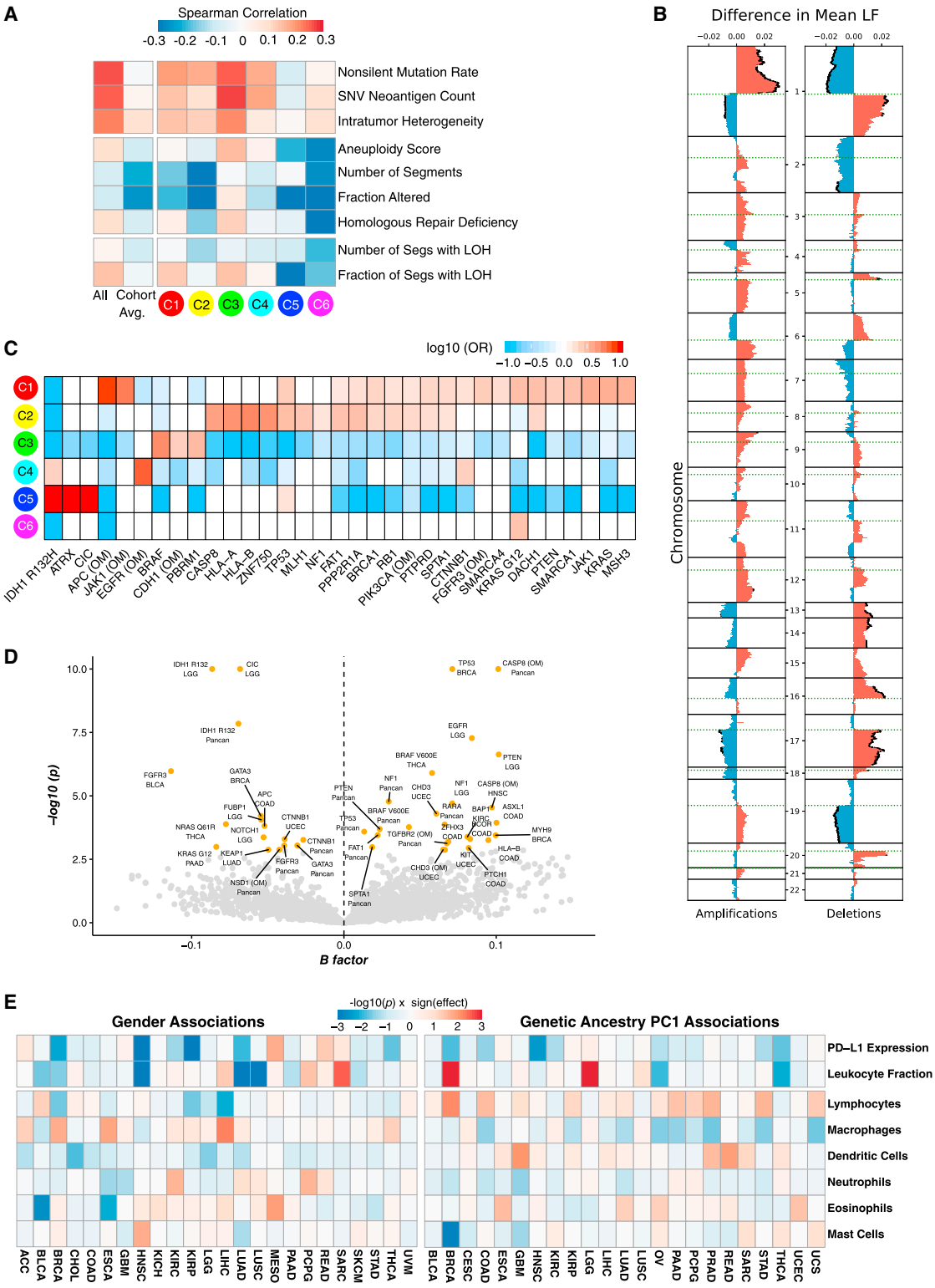


Figure 4. Immune Response and Genome State

(A) Correlation of DNA damage measures (rows) with LF. From left to right: all TCGA tumors; averaged over tumor type; grouped by immune subtype.

(B) LF association with copy number (CN) alterations. Left: Differences between observed and expected mean LF in tumors with amplifications, by genomic region. Significant (FDR < 0.01) differences in mean LF are marked with black caps on the profiles. Right: Same, for deletions.

(legend continued on next page)

focal copy number events versus larger events on immune infiltration in certain immune subtypes.

Specific SCNAs affected LF and immune composition (Figures 4B and S4A). Chromosome 1p (including *TNFRS9* and *VTCN1*) amplification associated with higher LF, while its deletion did the opposite. 19q deletion (including *TGFB1*) also correlated with lower LF, consistent with the role of TGF-β in immune cell recruitment (Bierie and Moses, 2010). Amplification of chr2, 20q, and 22q (including *CTLA4*, *CD40*, and *ADORA2*, respectively), and deletions of 5q, 9p, and chr19 (including *IL13* and *IL4*, *IFNA1* and *IFNA2*, and *ICAM1*, respectively) associated with changes in macrophage polarity (Figure S4A). IL-13 influences macrophage polarization (Mantovani et al., 2005), implying a possible basis for our observation that IL-13 deletions associated with altered M0 macrophage fractions.

Increased ITH associates with worse clinical outcomes or lower efficacy of IM therapy in a number of cancer types (McGranahan et al., 2016; Morris et al., 2016). ITH correlated (Spearman, Benjamini-Hochberg [BH]-adjusted $p < 0.05$) with total LF in nine tumor types (LUAD, BRCA, KIRC, HNSC, GBM [glioblastoma multiforme], OV, BLCA, SKCM, and READ; data not shown) and with individual relative immune cell fractions in many tumor types (Figure S4B). ITH was highest in C1 and C2 ($p < 10^{-229}$ relative to all others) and lowest in C3 ($p = 3 \times 10^{-5}$, Figure 1C), possibly supporting the link between lower ITH and improved survival.

We correlated mutations in 299 cancer driver genes with immune subtypes and found 33 significant associations ($q < 0.1$) (Figure 4C, Table S2). C1 was enriched in mutations in driver genes, such as *TP53*, *PIK3CA*, *PTEN*, or *KRAS*. C2 was enriched in many of these genes, as well as *HLA-A* and *B* and *CASP8*, which could be immune-evading mechanisms (Rooney et al., 2015). C3 was enriched in *BRAF*, *CDH1*, and *PBRM1* mutations, a finding of note since patients with *PBRM1* mutations respond particularly well to IM therapy (Miao et al., 2018). C4 was enriched in *CTNNB1*, *EGFR*, and *IDH1* mutations. C5 was enriched in *IDH1*, *ATRX*, and *CIC*, consistent with its predominance of LGG samples. C6 was only enriched in *KRAS* G12 mutations. Mutations in 23 driver genes associated with increased LF either in specific tumor types or across them, including *TP53*, *HLA-B*, *BRAF*, *PTEN*, *NF1*, *APC*, and *CASP8*. Twelve other events were associated with lower LF, including the *IDH1* R132H mutation, *GATA3*, *KRAS*, *NRAS*, *CTNNB1*, and *NOTCH1* (Figure 4D).

Since driver mutations in the same pathway had opposing correlations with LF (e.g., *BRAF*, *KRAS*, *NRAS*), we considered the overall effect of somatic alterations (mutations and SCNAs) on eight oncogenic signaling pathways. PI3K, NOTCH, and RTK/RAS pathway disruptions showed variable, tumor type-specific effects on immune factors, while TGF-β pathway disruptions more consistently associated with lower LF (most prominently in C2 and C6; Figure S4C), higher eosinophils (C2), and increased macrophages. However, in C3, TGF-β pathway disruption associated with higher LF and M1 macrophages and lower memory B cells, helper T cells, and M0 macrophages. Thus, TGF-β pathway disruption has context-dependent effects on LF but may promote increased macrophages, particularly M1. Higher M1/M2 ratio, in turn, may reiterate the local pro-inflammatory state in these patients.

Immune Response Correlates of Demographic and Germline Variation

Immune cell content and expression of *PD-L1* varied by gender and genetic ancestry (Figures 4E and S4D). *PD-L1* expression was greater ($p < 0.05$, Kruskal-Wallis test, unadjusted) in women than in men in HNSC, KIRC, LUAD, THCA, and KIRP (Figure S4E), while mesothelioma (MESO) showed an opposite trend. *PD-L1* expression was lower in individuals of predicted African ancestry (overall $p = 5 \times 10^{-6}$). This association was consistent across most cancer types and was significant ($p < 0.05$, unadjusted) in BRCA, COAD, HNSC (Figure S4F), and THCA. No single *cis*-eQTL significantly correlated with *PD-L1* expression, although the SNP rs822337, approximately 1 kb upstream of *CD274* transcription start, correlated weakly ($p = 0.074$; 1.3×10^{-4} unadjusted; Figure S4G). Lymphocyte fractions tended to be lower in people of Asian ancestry, particularly in UCEC (uterine corpus endometrial carcinoma) and BLCA (Figure S4H). The significance of these demographic associations remains unclear but provides hypotheses for the efficacy of checkpoint inhibitor therapy based on genetic ancestry.

Survey of Immunogenicity

Peptides predicted to bind with MHC proteins (pMHCs) and induce antitumor adaptive immunity were identified from SNV and indel mutations. The number of pMHCs (neoantigen load) varied between immune subtypes (Figure 1C), correlated positively with LF in most immune subtypes (Figure S4I), and trended positive in most TCGA tumor subtypes, with some negative correlation seen among GI subtypes, and differential trending seen among individual LUAD, LUSC, OV, and KIRP subtypes (Figure S4J). Neoantigen load also associated with higher content of CD8 T cells, M1 macrophages, and CD4 memory T cells, and lower Treg, mast, dendritic, and memory B cells in multiple tumor types (Figure S4K).

Most SNV-derived peptides which bind to MHC were each found in the context of a single MHC allele (89.9%). Single mutations generate 99.8% of unique pMHCs while 0.2% result from distinct mutations in different genetic loci yielding identical peptides (Figure 5A). The most frequently observed pMHCs

(C) Enrichment and depletion of mutations in driver genes and oncogenic mutations (OM) within immune subtypes, displayed as fold enrichment. Significance was evaluated by the Cochran-Mantel-Haenszel χ^2 test, to account for cancer type (white, no significant association).

(D) Volcano plot showing driver genes and OMs associated with changes in LF, across all tumors ("Pancan") and within specific tumor types as indicated. x axis: Multivariate correlation with LF (B-factor), taking into account tumor type and number of missense mutations. Values > 0 represent positive correlation with LF and vice versa; y axis: $-\log10(p)$. Significant events (FDR < 0.1; $p < 0.003$) are in orange, others in gray.

(E) Left: Degree of association between gender for eight selected immune characteristics (rows) within TCGA tumor types (columns). Blue denotes a higher value in women than in men, and red the opposite. Right: Degree of association between the immune characteristics and the first principal component of genetic ancestry in TCGA participants (PC1), reflecting degree of African ancestry. Blue reflects lower values in individuals of African descent.

See also Figure S4 and Table S2.

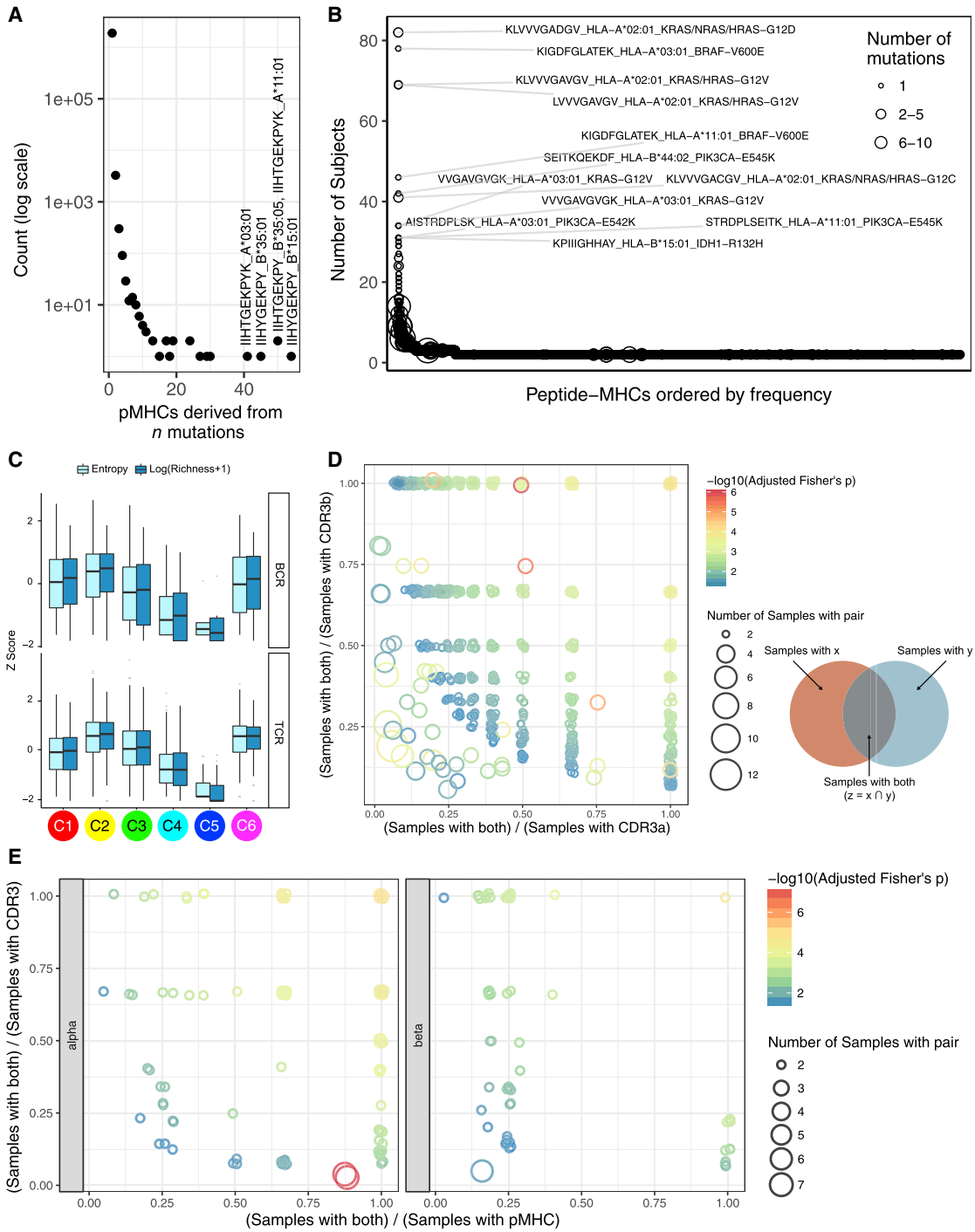


Figure 5. The Tumor-Immune Interface

(A) Distribution of the number of pMHCs associated with number of mutations; the 4 pMHCs derived from > 40 mutations are labeled.

(B) Numbers of tumors expressing shared pMHCs. The known cancer genes from which the most frequent pMHCs in the population are derived are indicated.

(C) BCR (top) and TCR (bottom) diversity measured by Shannon entropy and species richness, logarithmically transformed, and expressed as Z-scores, for immune subtypes.

(D and E) Co-occurrence of CDR3a-CDR3b (D) and pMHC-CDR3 pairs (E) as a surrogate marker for shared T cell responses. Pairs found in at least two samples and meeting statistical significance are plotted, with jitter. x and y axes indicate how exclusive the pair members are: pairs in the top right typically co-occur, whereas along the axes each member is more often found separately. Size of the circle indicates how many samples that pair was found in.

See also Figure S5 and Tables S3, S4, and S5.

were from recurrently mutated genes (*BRAF*, *IDH1*, *KRAS*, and *PIKC3A* for SNVs, *TP53* and *RNF43* for indels) (Figure 5B, Tables S3 and S4). In BRCA and LIHC, worse PFI was associated with higher neoantigen load, while BLCA and UCEC showed the opposite effect (Figure S5A). For most tumors, however, there were no clear associations between predicted pMHC count and survival. Within immune subtypes (Figure S5B), higher neoantigen load was associated with improved PFI in C1 and C2 and worse PFI in C3, C4, and C5. These results suggest that neoantigen load provides more prognostic information within immune subtypes than based on tissue of origin, emphasizing the importance of overall immune signaling in responding to tumor neoantigens.

Cancer testis antigens (CTA) overall expression, and that of individual CTAs, varied by immune subtype with C5 having the highest ($p < 10^{-13}$) and C3 the lowest ($p = 10^{-4}$) expression values (Figure 1C). *CEP55*, *TTK*, and *PBK* were broadly expressed across immune subtypes, with enrichment in C1 and C2. C5 demonstrated high expression of multiple CTAs, illustrating that CTA expression alone is insufficient to elicit an intratumoral immune response.

We found human papilloma virus (HPV) in 6.2% of cases, mainly in CESC, GBM, HNSC, and KIRC, whereas hepatitis B virus (HBV) and Epstein-Barr virus (EBV) were mainly found in LIHC and STAD (stomach adenocarcinoma), respectively. In a regression model of all tumors, high load of each virus type associated with immune features (Figure S5C, cancer-type adjusted). High EBV content associated strongly with high *CTLA4* and *CD274* expression and low B cell signatures. High HPV levels associated with increased proliferation and Th2 cells but low macrophage content. In contrast, high HBV levels associated with Th17 signal and γδ T cell content. These findings highlight the diverse effect of different viruses on the immune response in different cancer types.

Our findings suggest that pMHC burden and viral content impact immune cell composition, while CTAs have inconsistent effects on the immune response. Moreover, the effect of pMHC load on prognosis is disease specific and influenced by immune subtype.

The Adaptive Immune Receptor Repertoire in Cancer

Antigen-specific TCR and BCR repertoires are critical for recognition of pathogens and malignant cells and may reflect a robust anti-tumor response comprising a large number of antigen-specific adaptive immune cells that have undergone clonal expansion and effector differentiation.

We evaluated TCR α and β and immunoglobulin heavy and light chain repertoires from RNA-seq. Mean TCR diversity values differed by immune subtype, with the highest diversity in C6 and C2 ($p < 10^{-183}$, Wilcoxon, relative to all other subtypes; Figure 5C) and by tumor type (Figure S5D, lower panel). We saw recurrent TCR sequences across multiple samples (Figure S5E, Table S5), suggesting a common, but not necessarily cancer-related, antigen (the top recurrent TCRs include known mucosal associated invariant T cell sequences). We assessed co-occurrence of complementarity determining region 3 (CDR3) α and β chains, in order to determine the frequency of patients with identical TCRs (a surrogate marker for shared T cell responses). We identified 2,812 α-β pairs present in at least 2 tumors ($p \leq 0.05$, Fisher's exact test with Bonferroni correction; Figure 5D and Table S5). Likewise, testing for co-occurrence of specific SNV pMHC-CDR3 pairs across all patients identified 206 pMHC-CDR3 α pairs and 196 pMHC-CDR3 β pairs (Figure 5E, Table S5). Thus, a minority of these patients appear to share T cell responses, possibly mediated by public antigens. That said, there is relatively little pMHC and TCR sharing among tumors, highlighting the large degree of diversity in TILs.

Higher TCR diversity only correlated with improved PFI in a few tumor types (BLCA, COAD, LIHC, and UCEC) (Figure S5F). Therefore, it may be more important for the immune system to mount a robust response against only a few antigens, than a diverse response against many different antigens.

The pattern of immunoglobulin heavy chain diversity was similar to that of TCR diversity (Figures 5C and S5D), with tumors showing significant variance in IgH repertoire diversity, suggesting differential B cell recruitment and/or clonal expansion within the tumor types.

Regulation of Immunomodulators

IMs are critical for cancer immunotherapy with numerous IM agonists and antagonists being evaluated in clinical oncology (Tang et al., 2018). To advance this research, understanding of their expression and modes of control in different states of the TME is needed. We examined IM gene expression, SCNAs, and expression control via epigenetic and miRNA mechanisms.

Gene expression of IMs (Table S6, Figure 6A) varied across immune subtypes, and IM expression largely segregated tumors by immune subtypes (Figure S6A), perhaps indicative of their role in shaping the TME. Genes with the greatest differences between subtypes (Figures 6B and S6B) included *CXCL10* (BH-adjusted $p < 10^{-5}$), most highly expressed in C2 (consistent with its known interferon inducibility) and *EDNRB* (BH-adjusted $p < 10^{-5}$), most highly expressed in the immunologically quiet C5. DNA methylation of many IM genes, e.g., *CD40* (Figure 6C), *IL10*, and *IDO1*, inversely correlated with gene expression, suggesting epigenetic silencing. 294 miRNAs were implicated as possible regulators of IM gene expression; among these, several associated with IMs in multiple subtypes (Figure S6C) including immune inhibitors (*EDNRB*, *PD-L1*, and *VEGFA*) and activators (*CD28* and *TNFRSF9*). The immune activator *BTN3A1* was one of the most commonly co-regulated IMs from the SYGNAL-PanImmune network (below). Negative correlations between *miR-17* and *BTN3A1*, *PDCD1LG2*, and *CD274* may relate to the role of this miRNA in maturation and activation of cells into effector or memory subsets (Liang et al., 2015).

Copy-number alterations affected multiple IMs and varied across immune subtypes. C1 and C2 showed both frequent amplification and deletion of IM genes, consistent with their greater genomic instability, while subtypes C3 and C5 generally showed fewer alterations in IM genes. In particular, IMs *SLAMF7*, *SELP*, *TNFSF4* (*OX40L*), *IL10*, and *CD40* were amplified less frequently in C5 relative to all samples, while *TGFB1*, *KIR2DL1*, and *KIR2DL3* deletions were enriched in C5 (Figure 6D), consistent with our observation of lower immune infiltration with *TGFB1* deletion (Figure S4A). *CD40* was most frequently amplified in C1 (Figure 6D) (Fisher's exact $p < 10^{-10}$ for all comparisons mentioned). Overall, these marked differences in IM copy

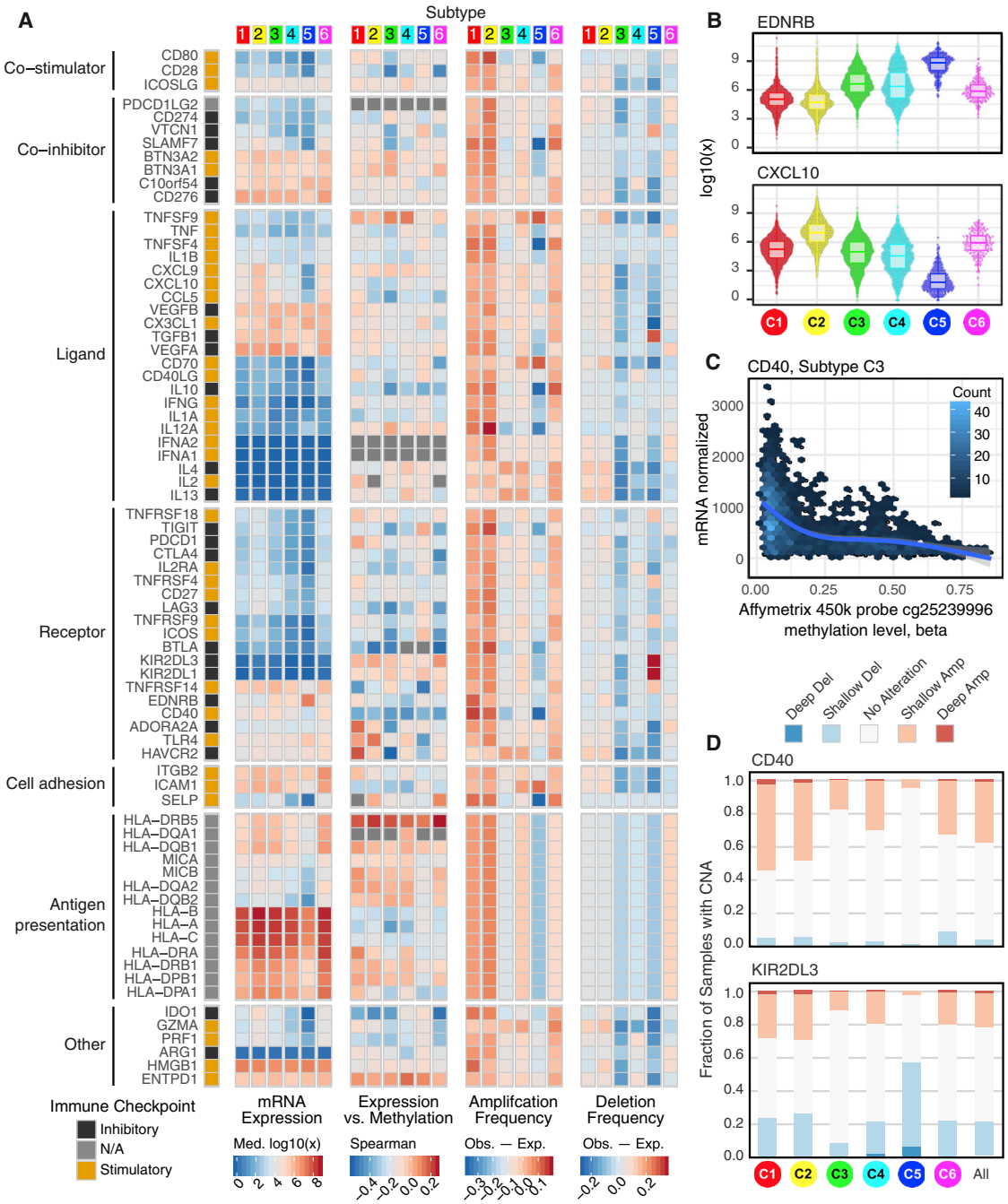


Figure 6. Regulation of Immunomodulators

(A) From left to right: mRNA expression (median normalized expression levels); expression versus methylation (gene expression correlation with DNA-methylation beta-value); amplification frequency (the difference between the fraction of samples in which an IM is amplified in a particular subtype and the amplification fraction in all samples); and the deletion frequency (as amplifications) for 75 IM genes by immune subtype.

(B) Distribution of log-transformed expression levels for IM genes with largest differences across subtypes (by Kruskal-Wallis test).

(C) *CD40* expression is inversely correlated to methylation levels (Affymetrix 450K probe cg25239996, 125 bases upstream of *CD40* TSS) in C3. Each point represents a tumor sample, and color indicates point density.

(D) Proportion of samples in each immune subtype with copy number alterations in *CD40* (top) and *KIR2DL3* (bottom). The "All" column shows the overall proportion (8,461 tumors).

See also Figure S6 and Table S6.

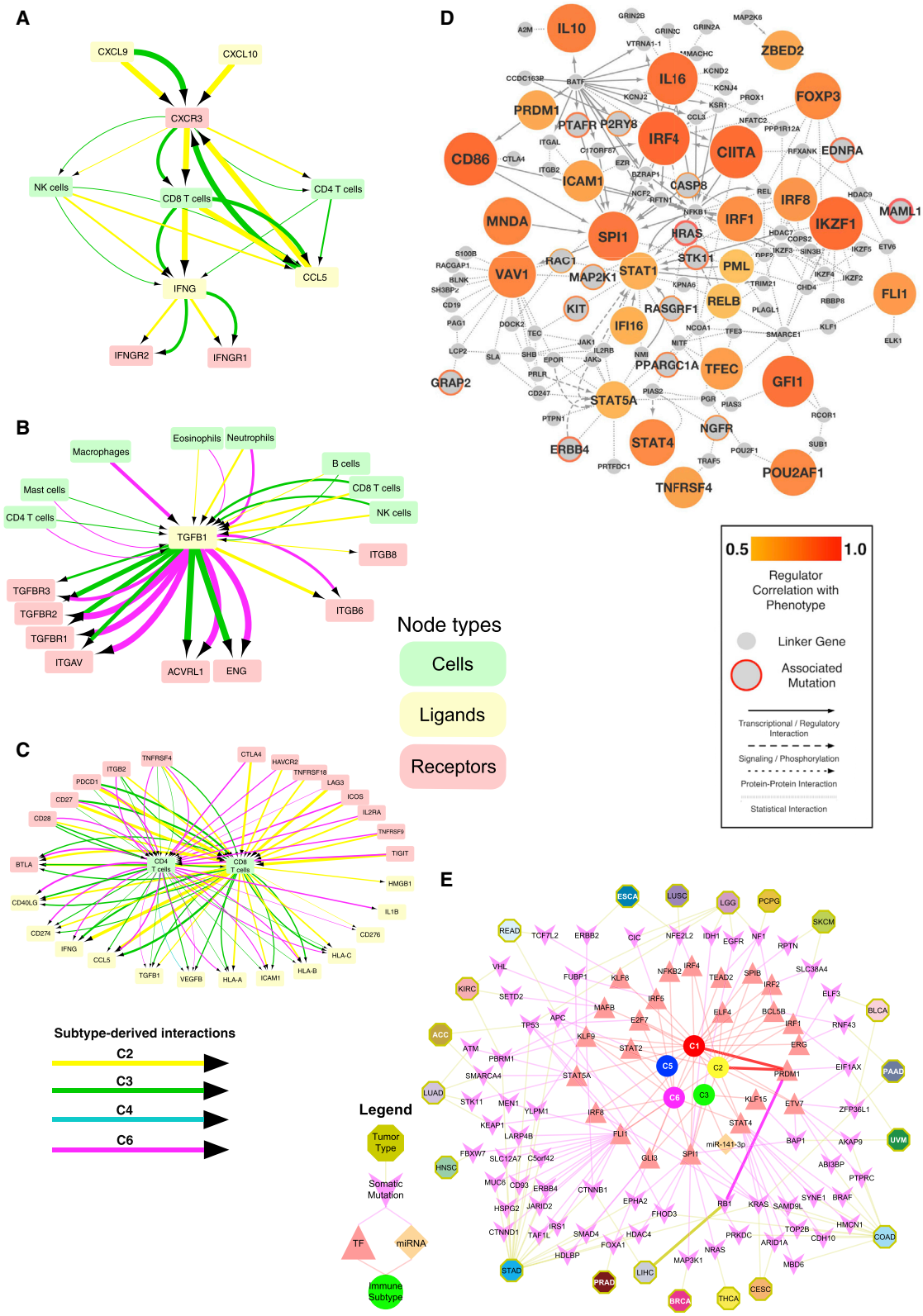


Figure 7. Predicted Networks Modulating the Immune Response to Tumors

TME estimates and tumor cell characteristics were combined with available data on possible physical, signaling, and regulatory interactions to predict cellular and molecular interactions involved in tumoral immune responses.

(legend continued on next page)

number may be reflective of more direct modulation of the TME by cancer cells.

Among IMs under investigation for cancer therapy, expression of *VISTA* is relatively high in all tumor types and highest in MESO; *BTLA* expression is high in C4 and C5; *HAVCR2* (*TIM-3*) shows evidence of differential silencing among immune subtypes; and *IDO1* is amplified, mostly in C1. The observed differences in regulation of IMs might have implications for therapeutic development and combination immune therapies, and the multiple mechanisms at play in evoking them further highlights their biological importance.

Networks Modulating Tumoral Immune Response

The immune response is determined by the collective states of *intracellular* molecular networks in tumor, immune, and other stromal cells and the *extracellular* network encompassing direct interaction among cells and communication via soluble proteins such as cytokines to mediate interactions among those cells.

Beginning with a large network of extracellular interactions known from other sources, we identified which of those met a specified precondition for interaction, namely that both interaction partners are consistently present within samples in an immune subtype, according to our TME estimates. We focused the network on IMs. Networks in C2 and C3 had abundant CD8 T cells, while C3, C4, and C6 were enriched in CD4 T cells.

A small sub-network (Figure 7A), focused around IFN-γ, illustrates some subtype-specific associations. In both C2 and C3, CD4 T cells, CD8 T cells, and NK cells correlated with expression of *IFNG* and *CCL5*, a potent chemoattractant. A second subnetwork (Figure 7B), centered on TGF-β, was found in the C2, C3, and C6 networks. Across subtypes, different cell types were associated with abundant expression of *TGFB1*: CD4 T cells and mast cells in C3 and C6, macrophages in C6, neutrophils and eosinophils in C2 and C6, and B cells, NK cells, and CD8 T cells in C2 and C3. The receptors known to bind TGF-β likewise were subtype specific and may help mediate the TGF-β-driven infiltrates, with *TGFBR1*, *2*, and *3* found only in the C3 and C6 networks. These results largely echo findings seen in our TGF-β pathway analysis (Figure S4C), which examined the effects of intracellular, rather than extracellular, signaling disruption on immune TME composition across immune subtypes. Finally, a third cytokine subnetwork illustrates variation in T cell ligands and receptors across immune subtypes (Figure 7C). CD4 and CD8 receptors fell into two groups, those found in C2, C3, and C6 networks, such as *PDCD1*, and those absent in C3, such as *IL2RA* and *LAG3*. Some T cell-associated ligands were subtype specific, such as *CD276* (C2, C6), *IL1B* (C6), and *VEGFB* (C4).

The derived extracellular networks reflect the properties of immune subtypes in terms of cellular propensities and immune pathway activation noted earlier (Figures 1B, 1C, 2A, and S2A), but also place those properties in the context of possible interactions in the TME that may play a role in sculpting those same properties. The particular associations observed among IMs within distinct subtypes may be important for identifying directions for therapy.

We next used two complementary approaches, master regulators (MRs) and SYGNAL, to synthesize a pan-cancer transcriptional regulatory network describing the interactions linking genomic events to transcriptional regulators to downstream target genes, and finally to immune infiltration and patient survival. In both approaches, somatic alterations were used as anchors to infer regulatory relationships, in that they can act as a root cause of the "downstream" transcriptional changes mediated through transcription factors (TFs) and miRNAs.

This resulted in two transcriptional networks. The first one, MR-PanImmune, consisted of 26 MRs that acted as hubs associated with observed gene expression and LF, connected with 15 putative upstream driver events (Figure 7D). The second

(A) Immune subtype-specific extracellular communication network involving IFN-γ (*IFNG*, bottom of the diagram), whose expression is concordant with that of its cognate receptors *IFNGR1* and *IFNGR2* (bottom right and left, respectively), in C2 and C3 (yellow and green arrows, respectively; line thickness indicates strength of association). NK cells (left), which are known to secrete IFN-γ, could be producing IFN-γ in C2 and C3, as the NK cellular fraction is concordant with *IFNG* expression in both. *CXCR3* is known to be expressed on NK cells and has concordant levels, but only in C3 (green arrow). This is a subnetwork within a larger network constructed by similarly combining annotations of known interactions between ligands, receptors, and particular immune cells types, with evidence for concordance of those components.

(B) TGF-β subnetwork. Magenta: C6.

(C) T cell subnetwork.

(D) Master Regulator (MR)-Pan-Immune Network. The network diagram shows 26 MR "hubs" (filled orange) significantly associated with 15 upstream driver events (orange rings), along with proteins linking the two. The lineage factor *VAV1* (on left) is inferred to be a MR by combining predicted protein activity with data on gene expression, protein interactions, and somatic alterations. VAV1 activity correlates with LF (degree of correlation depicted as degree of orange). Mutations in *HRAS* (center of network) are statistically associated with changes in LF. The HRAS and VAV1 proteins are in close proximity on a large network of known protein-protein interactions (not shown), as both can lead to activation of protein MAP2K1, (as shown connecting with dotted lines). Mutations in *HRAS* are associated ($p < 0.05$) with VAV1 activity, and their link through documented protein interactions implies that HRAS could directly modulate the activity of VAV1. In the diagram, the size of MR nodes represents their ranked activity. Smaller nodes with red borders represent mutated and/or copy-number altered genes statistically associated with one or more MR and LF, with the thickness of the border representing the number of associated MRs; small gray nodes are "linker" proteins.

(E) Regulators of immune subtypes from SYGNAL-PanImmune Network. Tumor types (octagons) linked through mutations (purple chevrons) to transcription factors (TFs, red triangles) and miRNAs (orange diamonds) that actively regulate the expression of IMs in biclusters associated with a single immune subtype (circles). The network describes predicted causal and mechanistic regulatory relationships linking tumor types through their somatic mutations (yellow edges) which causally modulate the activity of TFs and/or miRNAs (purple edges), which in turn regulate genes (not shown) whose expression is associated with an immune subtype (red edges). For example, *RB1* mutations in LIHC (5% of patients) have significant evidence for causally modulating the activity of PRDM1 which in turn regulates genes associated (causal model at least 3 times as likely as alternative models and p value < 0.05) with C1 and C2. Interactions for this path are bolded.

one, SYGNAL-Panimmune, comprised 171 biclusters enriched in IMs and associated with LF.

Seven TFs were shared between the MR- and SYGNAL-PanImmune networks, a significant overlap ($p = 4.8 \times 10^{-10}$, Fisher's exact test): *PRDM1*, *SPI1*, *FLI1*, *IRF4*, *IRF8*, *STAT4*, and *STAT5A*. Additional MRs included the hematopoietic lineage specific factor *IKZF1*, which may reflect variation in immune cell content, and known IMs, such as *IFNG*, *IL16*, *CD86*, and *TNFRSF4*. The regulators in SYGNAL-PanImmune were inferred to regulate a total of 27 IM genes (Figure S7C). The top two most commonly co-regulated IMs from SYGNAL-PanImmune, *BTN3A1* and *BTN3A2*, are of particular interest as they modulate the activation of T cells (Cubillos-Ruiz et al., 2010) and have antibody-based immunotherapies (Benyamine et al., 2016; Legut et al., 2015).

Somatic alterations in *AKAP9*, *HRAS*, *KRAS*, and *PREX2* were inferred to modulate the activity of IMs according to both the MR- and SYGNAL-PanImmune, a significant overlap ($p = 1.6 \times 10^{-7}$, Fisher's exact test). In MR-PanImmune, *MAML1* and *HRAS* had the highest number of statistical interactions with 26 MRs. This analysis identified complex roles for the RAS-signaling pathway (Figure 7D) specifically through connections to lineage factor *VAV1* (implicated in multiple human cancers), potentially mediated by *MAP2K1*. Similarly, *MAML1*, hypothesized to mediate cross-talk across pathways in cancer (McElhinny et al., 2008), was associated ($p \leq 0.05$) with multiple MRs, including *STAT1*, *STAT4*, *CIITA*, *SPI1*, *TNFRSF4*, *CD86*, *VAV1*, *IKZF1*, and *IL16.*

In SYGNAL-PanImmune, some regulators of IMs, but not upstream somatic mutations, were shared between tumor types, including *STAT4*, which regulated *BTN3A1* and *BTN3A2* in both LUSC and UCEC, secondary to implied causal mutations *TP53* and *ARHGAP35*, respectively. Conversely, causal mutations shared across tumor types may associate with different tumor-specific downstream regulators. *TP53* was a causal mutation in UCEC acting through *IRF7* to regulate many of the same IMs as was seen in LUSC. These differences in causal relationships arise because the different cell types giving rise to each tumor type affect oncogenic paths.

We identified the putative regulators of immune gene expression within immune subtypes (Figure 7E). In these predictions, C1-associated biclusters were regulated by *ERG*, *KLF8*, *MAFB*, *STAT5A*, and *TEAD2.* C1 and C2 shared regulation by *BCL5B*, *ETV7*, *IRF1*, *IRF2*, *IRF4*, *PRDM1*, and *SPIB*, consistent with IFN-γ signaling predominance in these subtypes. C3 was regulated by *KLF15* and *miR-141-3p*. C6-associated biclusters were regulated by *NFKB2*. C1, C2, and C6 shared regulation by *STAT2* and *STAT4*, implying shared regulation by important immune TF families, such as STAT and IRF, but also differential employment of subunits and family members by the immune milieu.

In SYGNAL-PanImmune, the increased expression of biclusters enriched with IMs from KIRC, LGG, LUSC, and READ was associated with worse patient survival (CoxPH BH adjusted p value ≤ 0.05). Conversely, the increased expression of biclusters enriched with IMs from SKCM, containing *CCL5*, *CXCL9*, *CXCL10*, *HAVCR2*, *PRF1*, and MHC class II genes, were associated with improved patient survival (BH-adjusted $p \leq 0.05$).

DISCUSSION

We report an extensive evaluation of immunogenomic features in more than 10,000 tumors from 33 cancer types. Data and results are available as Supplemental Tables, at NCI GDC, and interactively at the CRI iAtlas portal, which is being configured to accept new immunogenomics datasets and feature calculations as they come available, including those derived from immunotherapy clinical trials, to develop as a ''living resource'' for the immunogenomics community. Meta-analysis of consensus expression clustering revealed immune subtypes spanning multiple tumor types and characterized by a dominance of either macrophage or lymphocyte signatures, T-helper phenotype, extent of intratumoral heterogeneity, and proliferative activity. All tumor samples were assessed for immune content by multiple methods. These include the estimation of immune cell fractions from deconvolution of gene expression and DNA methylation data, prediction of neoantigen-MHC pairs from mutations and HLA-typing, and evaluation of BCR and TCR repertoire from RNA-sequencing data. Immune content was compared among immune and cancer subtypes, and somatic alterations were identified that correlate with changes in the TME. Finally, predictions were made of regulatory networks that could influence the TME, and intracellular communication networks in the TME, based on integrating known interactions and observed associations. Immunogenomic features were predictive of outcome, with OS and PFI differing between immune subtypes both within and across cancer types.

C4 and C6 subtypes conferred the worst prognosis on their constituent tumors and displayed composite signatures reflecting a macrophage dominated, low lymphocytic infiltrate, with high M2 macrophage content, consistent with an immunosuppressed TME for which a poor outcome would be expected. In contrast, tumors included in the two subtypes displaying a type I immune response, C2 and C3, had the most favorable prognosis, consistent with studies suggesting a dominant type I immune response is needed for cancer control (Galon et al., 2013). In addition, C3 demonstrated the most pronounced Th17 signature, in agreement with a recent systematic review suggesting that Th17 expression is generally associated with improved cancer survival (Punt et al., 2015). C2 was IFN-γ dominant and showed a less favorable survival despite having the highest lymphocytic infiltrate, a CD8 T cell-associated signature, and highest M1 content, suggesting a robust anti-tumor immune response. One explanation for this discrepancy is the aggressiveness of both the tumor types and specific cases within C2 relative to C3. C2 showed the highest proliferation signature and ITH while C3 was the lowest in both those categories. It may be that the immune response simply could not control the rapid growth of tumors comprising C2. A second hypothesis is that tumors in C2 are those that have already been remodeled by the existing robust type I infiltrate and have escaped immune recognition. While signatures biased toward interferon-mediated viral sensing and antigen presentation genes were often associated with higher survival, interferon signatures without increased antigen presentation showed an opposite association. Loss of genes associated with antigen processing and presentation is often found in tumors that have been immune edited. In contrast to the potential immune

editing of C2, C3 may represent immunologic control of disease, that is, immune equilibrium.

Possible impact of somatic alterations on immune response was seen. For example, *KRAS* mutations were enriched in C1 and but infrequent in C5, suggesting that mutations in driver oncogenes alter pathways that affect immune cells. Driver mutations such as *TP53*, by inducing genomic instability, may alter the immune landscape via the generation of neoantigens. Our findings confirmed previous work showing that mutations in *BRAF* (Ilieva et al., 2014) enhance the immune infiltrate while those in *IDH1* diminish it (Amankulor et al., 2017). Further work is needed to determine the functional aspects of these associations.

Tumor-specific neoantigens are thought to be key targets of anti-tumor immunity and are associated with improved OS and response to immune checkpoint inhibition in multiple tumor types (Brown et al., 2014). We found OS correlated with pMHC number in only a limited number of tumors, with no clear association in most tumors, including several responsive to immune checkpoint inhibitor therapy. There are some caveats to this finding. The current predictors are highly sensitive but poorly specific for neoantigen identification, and our approach did not include neoantigens from introns or spliced variants. Moreover, it is not possible to fully determine the ability to process and present an epitope or the specific T cell repertoire in each tumor, which impacts the ability to generate a neoantigen response. It is also possible that the role of neoantigens may vary with tumor type, as supported by our per-tumor results.

Integrative methods predicted tumor-intrinsic and tumor-extrinsic regulation in, of, and by the TME and yielded information on specific modes of intracellular and extracellular control, the latter reflecting the network of cellular communication among immune cells in the TME. The resulting network was rich in structure, with mast cells, neutrophils, CD4 T cells, NK cells, B cells, eosinophils, macrophages, and CD8 T cells figuring prominently. The cellular communication network highlighted the role of key receptor and ligands such as *TGFB1*, *CXCL10*, and *CXCR3* and receptor-ligand pairs, such as the *CCL5-CCR5* axis, and illustrated how immune cell interactions may differ depending on the immune system context, manifested in the immune subtype.

Predicted intracellular networks implied that seven immune-related TFs (including interferon and STAT-family transcription factors) may play an active role in transcriptional events related to leukocyte infiltration, and that mutations in six genes (including Ras-family proteins) may influence immune infiltration. Across tumor types, the TFs and miRNAs regulating the expression of IMs tended to be shared, while somatic mutations modulating those regulatory factors tended to differ. This suggests that therapies targeting regulatory factors upstream of IMs should be considered and that they may have a broader impact across tumor types than therapies focusing on somatic mutations. Of note, in these approaches, it is not always possible to fully ascertain whether some particular interaction acts in the tumor, immune, or stromal cell compartments, but this could be improved on by incorporating additional cell-type-specific knowledge. Shared elements of intra- and extracellular network models should also be explored, with particular regard to the IMs and cytokines in both.

There are important caveats to using TCGA data. First, survival event rates and follow-up durations differ across the tumor types. Second, for most tumor types, samples with less than 60% tumor cell nuclei by pathologist review were excluded from study, thus potentially removing the most immune-infiltrated tumors from analysis. The degree to which this biases the results, relative to the general population of cancer patients, is difficult to ascertain. Our analyses were also limited by restriction to data from genome-wide molecular assays, in the absence of targeted classical cellular immunology assays for confirming cell phenotype distribution, as those types of data have not been collected from TCGA patients.

In summary, six stable and reproducible immune subtypes were found to encompass nearly all human malignancies. These subtypes were associated with prognosis, genetic, and immune modulatory alterations that may shape the specific types of immune environments we have observed. With our increasing understanding that the tumor immune environment plays an important role in prognosis as well as response to therapy, the definition of the immune subtype of a tumor may play a critical role in the predicting disease outcome as opposed to relying solely on features specific to individual cancer types.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human Subjects
  - Sample Inclusion Criteria
- METHOD DETAILS
  - Clinical and Molecular Data
  - Immune Subtype Identification
  - Leukocyte and Stromal Fractions
  - Immune Cellular Fraction Estimates
  - Prognostic Correlations of Immune Phenotypes
  - Copy Number and DNA Damage Scores
  - Genomic Correlations with Immune Phenotypes
  - Genetic Ancestry
  - Identification of Neoantigens
  - Genomic Viral Content Analysis
  - T- and B- Cell Receptor Analysis
  - Immunomodulator Identification and Analysis
  - The Cell-to-Cell Communication Network
  - Master Regulators of Immune Genes
  - SYstems Genetics Network AnaLysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- SOFTWARE AND DATA AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and six tables and can be found with this article online at https://doi.org/10.1016/j.immuni.2018.03.023.

ACKNOWLEDGMENTS

We are grateful to all the patients and families who contributed to this study. We also thank the Office of Cancer Genomics at the NCI for organizational and logistical support of this study. The high-throughput analyses in this study were performed on the Institute for Systems Biology-Cancer Genomics Cloud (ISB-CGC) under contract number HHSN261201400007C and on the Seven Bridges Cancer Genomics Cloud under contract HHSN261201400008C, with federal funds from the National Cancer Institute, NIH, Department of Health and Human Services. Funding from the Cancer Research Institute is gratefully acknowledged, as is support from NCI through U54 HG003273, U54 HG003067, U54 HG003079, U24 CA143799, U24 CA143835, U24 CA143840, U24 CA143843, U24 CA143845, U24 CA143848, U24 CA143858, U24 CA143866, U24 CA143867, U24 CA143882, U24 CA143883, U24 CA144025, and P30 CA016672. The study was supported by W81XWH-12-2-0050, HU0001-16-2-0004 from the US Department of Defense through the Henry M. Jackson Foundation for the Advancement of Military Medicine. We thank Peter Hammerman and Yasin Şenbabaoğlu for contributions in early phases of this work.

AUTHOR CONTRIBUTIONS

Analysis, Computation, and Software: V.T., D.L.G., S.D.B., D.W., D.S.B., T.O.Y., E.P.-P., G.F.G., C.L.P., J.A.E., E.Z., A.C.C., E.O.P., I.K.A.S., A.J.G., R.M., F.F., A. Colaprico, J.S.P., L.E.M., N.S.V., J.L., Y.L., V.D., S.M.R., R.B., A.D.C., D.B., A.R., A.K., H.H., T.M.M., H.N., C.S.P., S.B., A.I.O., A.L., W.Z., J.G., J.S., B.G.V. Supervision: J.R., A. Califano, D.A., K.C., H.S., T.K.C., J.N.W., J.G., R.A.H., B.G.V., I.S. Writing: V.T., D.L.G., S.D.B., D.W., D.S.B., T.O.Y., E.P.-P., G.F.G., C.L.P., E.Z., A.C.C., E.O.P., I.K.A.S., A.J.G., R.M., F.F., A. Colaprico, N.S.V., H.H., T.M.M., H.N., J.S., C.E.R., A.J.L., J.S.S., E.G.D., M.L.D., B.G.V., I.S.

DECLARATION OF INTERESTS

Michael Seiler, Peter G. Smith, Ping Zhu, Silvia Buonamici, and Lihua Yu are employees of H3 Biomedicine, Inc. Parts of this work are the subject of a patent application: WO2017040526 titled "Splice variants associated with neomorphic sf3b1 mutants." Shouyoung Peng, Anant A. Agrawal, James Palacino, and Teng Teng are employees of H3 Biomedicine, Inc. Andrew D. Cherniack, Ashton C. Berger, and Galen F. Gao receive research support from Bayer Pharmaceuticals. Gordon B. Mills serves on the External Scientific Review Board of Astrazeneca. Anil Sood is on the Scientific Advisory Board for Kiyatec and is a shareholder in BioPath. Jonathan S. Serody receives funding from Merck, Inc. Kyle R. Covington is an employee of Castle Biosciences, Inc. Preethi H. Gunaratne is founder, CSO, and shareholder of NextmiRNA Therapeutics. Christina Yau is a part-time employee/consultant at NantOmics. Franz X. Schaub is an employee and shareholder of SEngine Precision Medicine, Inc. Carla Grandori is an employee, founder, and shareholder of SEngine Precision Medicine, Inc. Robert N. Eisenman is a member of the Scientific Advisory Boards and shareholder of Shenogen Pharma and Kronos Bio. Daniel J. Weisenberger is a consultant for Zymo Research Corporation. Joshua M. Stuart is the founder of Five3 Genomics and shareholder of NantOmics. Marc T. Goodman receives research support from Merck, Inc. Andrew J. Gentles is a consultant for Cibermed. Charles M. Perou is an equity stock holder, consultant, and Board of Directors member of BioClassifier and GeneCentric Diagnostics and is also listed as an inventor on patent applications on the Breast PAM50 and Lung Cancer Subtyping assays. Matthew Meyerson receives research support from Bayer Pharmaceuticals; is an equity holder in, consultant for, and Scientific Advisory Board chair for OrigiMed; and is an inventor of a patent for EGFR mutation diagnosis in lung cancer, licensed to LabCorp. Eduard Porta-Pardo is an inventor of a patent for domainXplorer. Han Liang is a shareholder and scientific advisor of Precision Scientific and Eagle Nebula. Da Yang is an inventor on a pending patent application describing the use of antisense oligonucleotides against specific lncRNA sequence as diagnostic and therapeutic tools. Yonghong Xiao was an employee and shareholder of TESARO, Inc. Bin Feng is an employee and shareholder of TESARO, Inc. Carter Van Waes received research funding for the study of IAP inhibitor ASTX660 through a Cooperative Agreement between NIDCD, NIH, and Astex Pharmaceuticals. Raunaq Malhotra is an employee and shareholder of Seven Bridges, Inc. Peter W. Laird serves on the Scientific Advisory Board for AnchorDx. Joel Tepper is a consultant at EMD Serono. Kenneth Wang serves on the Advisory Board for Boston Scientific, Microtech, and Olympus. Andrea Califano is a founder, shareholder, and advisory board member of DarwinHealth, Inc. and a shareholder and advisory board member of Tempus, Inc. Toni K. Choueiri serves as needed on advisory boards for Bristol-Myers Squibb, Merck, and Roche. Lawrence Kwong receives research support from Array BioPharma. Sharon E. Plon is a member of the Scientific Advisory Board for Baylor Genetics Laboratory. Beth Y. Karlan serves on the Advisory Board of Invitae.

Received: July 21, 2017
Revised: January 23, 2018
Accepted: March 21, 2018
Published: April 5, 2018

REFERENCES

Agarwal, V., Bell, G.W., Nam, J.W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. eLife *4*, https://doi.org/10.7554/eLife.05005.

Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H., and Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nat. Genet. *48*, 838–847.

Amankulor, N.M., Kim, Y., Arora, S., Kargl, J., Szulzewsky, F., Hanke, M., Margineantu, D.H., Rao, A., Bolouri, H., Delrow, J., et al. (2017). Mutant IDH1 regulates the tumor-associated immune system in gliomas. Genes Dev. *31*, 774–786.

Aten, J.E., Fuller, T.F., Lusis, A.J., and Horvath, S. (2008). Using genetic markers to orient the edges in quantitative trait networks: the NEO software. BMC Syst. Biol. *2*, 34.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202-208.

Bailey, P., Chang, D.K., Nones, K., Johns, A.L., Patch, A.M., Gingras, M.C., Miller, D.K., Christ, A.N., Bruxner, T.J., Quinn, M.C., et al. (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. Nature *531*, 47–52.

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108–112.

Beck, A.H., Espinosa, I., Edris, B., Li, R., Montgomery, K., Zhu, S., Varma, S., Marinelli, R.J., van de Rijn, M., and West, R.B. (2009). The macrophage colony-stimulating factor 1 response signature in breast carcinoma. Clin. Cancer Res. *15*, 778–787.

Bedognetti, D., Hendrickx, W., Ceccarelli, M., Miller, L.D., and Seliger, B. (2016). Disentangling the relationship between tumor genetic programs and immune responsiveness. Curr. Opin. Immunol. *39*, 150–158.

Benyamine, A., Le Roy, A., Mamessier, E., Gertner-Dardenne, J., Castanier, C., Orlanducci, F., Pouyet, L., Goubard, A., Collette, Y., Vey, N., et al. (2016). BTN3A molecules considerably improve Vgamma9Vdelta2T cells-based immunotherapy in acute myeloid leukemia. OncoImmunology *5*, e1146843.

Bierie, B., and Moses, H.L. (2010). Transforming growth factor beta (TGF-beta) and inflammation in cancer. Cytokine Growth Factor Rev. *21*, 49–59.

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A.C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity *39*, 782–795.

Bolotin, D.A., Shugay, M., Mamedov, I.Z., Putintseva, E.V., Turchaninova, M.A., Zvyagin, I.V., Britanova, O.V., and Chudakov, D.M. (2013). MiTCR: software for T-cell receptor sequencing data analysis. Nat. Methods *10*, 813–814.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Brown, S.D., Warren, R.L., Gibb, E.A., Martin, S.D., Spinelli, J.J., Nelson, B.H., and Holt, R.A. (2014). Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. Genome Res. *24*, 743–750.

Brown, S.D., Raeburn, L.A., and Holt, R.A. (2015). Profiling tissue-resident T cell repertoires by RNA sequencing. Genome Med. *7*, 125.

Calabro, A., Beissbarth, T., Kuner, R., Stojanov, M., Benner, A., Asslaber, M., Ploner, F., Zatloukal, K., Samonigg, H., Poustka, A., et al. (2009). Effects of infiltrating lymphocytes and estrogen receptor on gene expression and prognosis in breast cancer. Breast Cancer Res. Treat. *116*, 69–77.

Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. *30*, 413–421.

Ceccarelli, M., Barthel, F.P., Malta, T.M., Sabedot, T.S., Salama, S.R., Murray, B.A., Morozova, O., Newton, Y., Radenbaugh, A., Pagnotta, S.M., et al. (2016). Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. Cell *164*, 550–563.

Chang, H.Y., Sneddon, J.B., Alizadeh, A.A., Sood, R., West, R.B., Montgomery, K., Chi, J.T., van de Rijn, M., Botstein, D., and Brown, P.O. (2004). Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS Biol. *2*, E7.

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., Hackl, H., and Trajanoski, Z. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype pelationships and predictors of response to checkpoint blockade. Cell Rep. *18*, 248–262.

Chen, J.C., Alvarez, M.J., Talos, F., Dhruv, H., Rieckhof, G.E., Iyer, A., Diefes, K.L., Aldape, K., Berens, M., Shen, M.M., et al. (2014). Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. Cell *159*, 402–414.

Cheng, W.Y., Ou Yang, T.H., and Anastassiou, D. (2013a). Biomolecular events in cancer revealed by attractor metagenes. PLoS Comput. Biol. *9*, e1002920.

Cheng, W.Y., Ou Yang, T.H., and Anastassiou, D. (2013b). Development of a prognostic model for breast cancer survival in an open challenge environment. Sci. Transl. Med. *5*, 181ra150.

Chu, J., Sadeghi, S., Raymond, A., Jackman, S.D., Nip, K.M., Mar, R., Mohamadi, H., Butterfield, Y.S., Robertson, A.G., and Birol, I. (2014). BioBloom tools: fast, accurate and memory-efficient host species sequence screening using bloom filters. Bioinformatics *30*, 3402–3404.

Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I., et al. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. *44*, e71.

Cubillos-Ruiz, J.R., Martinez, D., Scarlett, U.K., Rutkowski, M.R., Nesbeth, Y.C., Camposeco-Jacobs, A.L., and Conejo-Garcia, J.R. (2010). CD277 is a negative co-stimulatory molecule universally expressed by ovarian cancer microenvironmental cells. Oncotarget *1*, 329–338.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Drake, J.M., Paull, E.O., Graham, N.A., Lee, J.K., Smith, B.A., Titz, B., Stoyanova, T., Faltermeier, C.M., Uzunangelov, V., Carlin, D.E., et al. (2016). Phosphoproteome integration reveals patient-specific networks in prostate cancer. Cell *166*, 1041–1054.

Ellrott, K., Bailey, M.H., Saksena, G., Covington, K.R., Kandoth, C., Stewart, C., Hess, J., Ma, S., McLellan, M., Sofia, H.J., et al. (2018). Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. Cell Syst. *6*, https://doi.org/10.1016/j.cels.2018.03.002.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. *33*, 1–22.

Galon, J., Angell, H.K., Bedognetti, D., and Marincola, F.M. (2013). The continuum of cancer immunosurveillance: prognostic, predictive, and mechanistic signatures. Immunity *39*, 11–26.

Gentles, A.J., Newman, A.M., Liu, C.L., Bratman, S.V., Feng, W., Kim, D., Nair, V.S., Xu, Y., Khuong, A., Hoang, C.D., et al. (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat. Med. *21*, 938–945.

Godec, J., Tan, Y., Liberzon, A., Tamayo, P., Bhattacharya, S., Butte, A.J., Mesirov, J.P., and Haining, W.N. (2016). Compendium of immune signatures identifies conserved and species-specific biology in response to inflammation. Immunity *44*, 194–206.

Gusenleitner, D., Howe, E.A., Bentink, S., Quackenbush, J., and Culhane, A.C. (2012). iBBiG: iterative binary bi-clustering of gene sets. Bioinformatics *28*, 2484–2492.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics *14*, 7.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

Hendrickx, W., Simeone, I., Anjum, S., Mokrab, Y., Bertucci, F., Finetti, P., Curigliano, G., Seliger, B., Cerulo, L., Tomei, S., et al. (2017). Identification of genetic determinants of breast cancer immune phenotypes by integrative genome-scale analysis. OncoImmunology *6*, e1253654.

Hornik, K. (2005). A CLUE for CLUster ensembles. J. Stat. Softw. *14*, 1–25.

Hugo, W., Zaretsky, J.M., Sun, L., Song, C., Moreno, B.H., Hu-Lieskovan, S., Berent-Maoz, B., Pang, J., Chmielowski, B., Cherry, G., et al. (2016). Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. Cell *165*, 35–44.

Hundal, J., Carreno, B.M., Petti, A.A., Linette, G.P., Griffith, O.L., Mardis, E.R., and Griffith, M. (2016). pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. Genome Med. *8*, 11.

Iglesia, M.D., Parker, J.S., Hoadley, K.A., Serody, J.S., Perou, C.M., and Vincent, B.G. (2016). Genomic analysis of immune cell infiltrates across 11 tumor types. J. Natl. Cancer Inst. *108*, https://doi.org/10.1093/jnci/djw144.

Ilieva, K.M., Correa, I., Josephs, D.H., Karagiannis, P., Egbuniwe, I.U., Cafferkey, M.J., Spicer, J.F., Harries, M., Nestle, F.O., Lacy, K.E., et al. (2014). Effects of BRAF mutations and BRAF inhibition on immune responses to melanoma. Mol. Cancer Ther. *13*, 2769–2783.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007). The role of site accessibility in microRNA target recognition. Nat. Genet. *39*, 1278–1284.

Khurana, E., Fu, Y., Chen, J., and Gerstein, M. (2013). Interpretation of genomic variants using a unified biological network approach. PLoS Comput. Biol. *9*, e1002886.

Knijnenburg, T., Wang, L., Zimmermann, M., Chambwe, N., Gao, G., Cherniack, A., Fan, H., Shen, H., Way, G., Greene, C., et al. (2018). Genomic and molecular landscape of DNA damage repair deficiency across The Cancer Genome Atlas. Cell Rep. *23*, https://doi.org/10.1016/j.celrep.2018.03.076.

Langfelder, P., and Horvath, S. (2007). Eigengene networks for studying the relationships between co-expression modules. BMC Syst. Biol. *1*, 54.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Lefranc, M.P., Giudicelli, V., Ginestoux, C., Jabado-Michaloud, J., Folch, G., Bellahcene, F., Wu, Y., Gemrot, E., Brochet, X., Lane, J., et al. (2009). IMGT, the international ImMunoGeneTics information system. Nucleic Acids Res. *37*, D1006–D1012.

Legut, M., Cole, D.K., and Sewell, A.K. (2015). The promise of gammadelta T cells and the gammadelta T cell receptor for cancer immunotherapy. Cell. Mol. Immunol. *12*, 656–668.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, M.X., Yeung, J.M., Cherny, S.S., and Sham, P.C. (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Hum. Genet. *131*, 747–756.

Li, B., Severson, E., Pignon, J.C., Zhao, H., Li, T., Novak, J., Jiang, P., Shen, H., Aster, J.C., Rodig, S., et al. (2016). Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol. *17*, 174.

Liang, Y., Pan, H.F., and Ye, D.Q. (2015). microRNAs function in CD8+T cell biology. J. Leukoc. Biol. *97*, 487–497.

Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to drive high quality survival outcome analytics. Cell *173*, https://doi.org/10.1016/j.cell.2018.02.052.

Mantovani, A., Sica, A., and Locati, M. (2005). Macrophage polarization comes of age. Immunity *23*, 344–346.

Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics *7* (*Suppl 1*), S7.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283.

McElhinny, A.S., Li, J.L., and Wu, L. (2008). Mastermind-like transcriptional coactivators: emerging roles in regulating cross talk among multiple signaling pathways. Oncogene *27*, 5138–5147.

McGranahan, N., Furness, A.J., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S.K., Jamal-Hanjani, M., Wilson, G.A., Birkbak, N.J., Hiley, C.T., et al. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science *351*, 1463–1469.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122.

Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. *12*, R41.

Miao, D., Margolis, C.A., Gao, W., Voss, M.H., Li, W., Martini, D.J., Norton, C., Bosse, D., Wankowicz, S.M., Cullen, D., et al. (2018). Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. Science *359*, 801–806.

Morris, L.G., Riaz, N., Desrichard, A., Senbabaoglu, Y., Hakimi, A.A., Makarov, V., Reis-Filho, J.S., and Chan, T.A. (2016). Pan-cancer analysis of intratumor heterogeneity as a prognostic determinant of survival. Oncotarget *7*, 10051–10063.

Mose, L.E., Selitsky, S.R., Bixby, L.M., Marron, D.L., Iglesia, M.D., Serody, J.S., Perou, C.M., Vincent, B.G., and Parker, J.S. (2016). Assembly-based inference of B-cell receptor repertoires from short read RNA sequencing data with V'DJer. Bioinformatics *32*, 3729–3734.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457.

Nielsen, M., and Andreatta, M. (2016). NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. Genome Med. *8*, 33.

Paull, E.O., Carlin, D.E., Niepel, M., Sorger, P.K., Haussler, D., and Stuart, J.M. (2013). Discovering causal pathways linking genomic events to transcriptional states using Tied Diffusion Through Interacting Events (TieDIE). Bioinformatics *29*, 2757–2764.

Pavesi, G., and Pesole, G. (2006). Using Weeder for the discovery of conserved transcription factor binding sites. Curr Protoc Bioinformatics *Chapter 2*. Unit 2 11.

Pencina, M.J., and D'Agostino, R.B. (2004). Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Stat. Med. *23*, 2109–2123.

Plaisier, C.L., Horvath, S., Huertas-Vazquez, A., Cruz-Bautista, I., Herrera, M.F., Tusie-Luna, T., Aguilar-Salinas, C., and Pajukanta, P. (2009). A systems genetics approach implicates USF1, FADS3, and other causal candidate genes for familial combined hyperlipidemia. PLoS Genet. *5*, e1000642.

Plaisier, C.L., Pan, M., and Baliga, N.S. (2012). A miRNA-regulatory network explains how dysregulated miRNAs perturb oncogenic processes across diverse cancers. Genome Res. *22*, 2302–2314.

Plaisier, C.L., O'Brien, S., Bernard, B., Reynolds, S., Simon, Z., Toledo, C.M., Ding, Y., Reiss, D.J., Paddison, P.J., and Baliga, N.S. (2016). Causal mechanistic regulatory network for glioblastoma deciphered using systems genetics network analysis. Cell Syst. *3*, 172–186.

Porta-Pardo, E., and Godzik, A. (2016). Mutation drivers of immunological responses to cancer. Cancer Immunol. Res. *4*, 789–798.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.

Punt, S., Langenhoff, J.M., Putter, H., Fleuren, G.J., Gorter, A., and Jordanova, E.S. (2015). The correlations between IL-17 vs. Th17 cells and cancer patient survival: a systematic review. OncoImmunology *4*, e984547.

Ramilowski, J.A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V.P., Itoh, M., Kawaji, H., Carninci, P., Rost, B., et al. (2015). A draft network of ligand-receptor-mediated multicellular signalling in human. Nat. Commun. *6*, 7866.

Reiss, D.J., Plaisier, C.L., Wu, W.J., and Baliga, N.S. (2015). cMonkey2: Automated, systematic, integrated detection of co-regulated gene modules for any organism. Nucleic Acids Res. *43*, e87.

Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell *160*, 48–61.

Saltz, J.H., Gupta, R., Hou, L., Kurc, T., Singh, P., Nguyen, V., Samaras, D., Shroyer, K.R., Zhao, T., Batiste, R., et al. (2018). Spatial organization and molecular correlation of tumor infiltrating lymphocytes using deep learning on pathology images. Cell Rep. *23*, https://doi.org/10.1016/j.celrep.2018.03.086.

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W.K., Luna, A., La, K., Dimitriadoy, S., Liu, D.L., Kantheti, H.S., Saghafinia, S., et al. (2018). Oncogenic signaling pathways in The Cancer Genome Atlas. Cell *173*.

Scrucca, L., Fop, M., Murphy, T.B., and Raftery, A.E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. R J. *8*, 289–317.

Senbabaoglu, Y., Gejman, R.S., Winer, A.G., Liu, M., Van Allen, E.M., de Velasco, G., Miao, D., Ostrovnaya, I., Drill, E., Luna, A., et al. (2016). Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. Genome Biol. *17*, 231.

Shukla, S.A., Rooney, M.S., Rajasagi, M., Tiao, G., Dixon, P.M., Lawrence, M.S., Stevens, J., Lane, W.J., Dellagatta, J.L., Steelman, S., et al. (2015). Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat. Biotechnol. *33*, 1152–1158.

Silva, T.C., Colaprico, A., Olsen, C., D'Angelo, F., Bontempi, G., Ceccarelli, M., and Noushmehr, H. (2016). TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages. F1000Res. *5*, 1542.

Siragusa, E., Weese, D., and Reinert, K. (2013). Fast and accurate read mapping with approximate seeds and multiple backtracking. Nucleic Acids Res. *41*, e78.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Szolek, A., Schubert, B., Mohr, C., Sturm, M., Feldhahn, M., and Kohlbacher, O. (2014). OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics *30*, 3310–3316.

Tang, J., Shalabi, A., and Hubbard-Lucey, V.M. (2018). Comprehensive analysis of the clinical immuno-oncology landscape. Ann. Oncol. *29*, 84–91.

Tatlow, P.J., and Piccolo, S.R. (2016). A cloud-based workflow to quantify transcript-expression levels in public cancer compendia. Sci. Rep. *6*, 39259.

Taylor, A.M., Shih, J., Ha, G., Gao, G.F., Zhang, X., Berger, A.C., Schumacher, S.E., Wang, C., Hu, H., Liu, J., et al. (2018). Genomic and functional approaches to understanding cancer aneuploidy. Cancer Cell *33*, https://doi.org/10.1016/j.ccell.2018.03.007.

Teschendorff, A.E., Gomez, S., Arenas, A., El-Ashry, D., Schmidt, M., Gehrmann, M., and Caldas, C. (2010). Improved prognostic classification of breast cancer defined by antagonistic activation patterns of immune response pathway modules. BMC Cancer *10*, 604.

The Cancer Genome Atlas Network (2015). Genomic classification of cutaneous melanoma. Cell *161*, 1681–1696.

The Cancer Genome Atlas Research Network (2011). Integrated genomic analyses of ovarian carcinoma. Nature *474*, 609–615.

Tibshirani, R., and Walther, G. (2005). Cluster validation by prediction strength. J. Comput. Graph. Stat. *14*, 511–528.

Venteicher, A.S., Tirosh, I., Hebert, C., Yizhak, K., Neftel, C., Filbin, M.G., Hovestadt, V., Escalante, L.E., Shaw, M.L., Rodman, C., et al. (2017). Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. Science *355*, https://doi.org/10.1126/science.aai8478.

Wingender, E., Schoeps, T., and Donitz, J. (2013). TFClass: an expandable hierarchical classification of human transcription factors. Nucleic Acids Res. *41*, D165–D170.

Wolf, D.M., Lenburg, M.E., Yau, C., Boudreau, A., and van 't Veer, L.J. (2014). Gene co-expression modules as clinically relevant hallmarks of breast cancer diversity. PLoS ONE *9*, e88309.

Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B., Lawrence, M.S., Zhsng, C.Z., Wala, J., Mermel, C.H., et al. (2013). Pan-cancer patterns of somatic copy number alteration. Nat. Genet. *45*, 1134–1140.

Zhang, Q.C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C.A., Bisikirska, B., Lefebvre, C., Accili, D., Hunter, T., et al. (2012). Structure-based prediction of protein-protein interactions on a genome-wide scale. Nature *490*, 556–560.